School of Doctoral Studies in Biological Sciences

University of South Bohemia in České Budějovice

Faculty of Science

Drivers of karyotype evolution in Lepidoptera

Ph.D. Thesis

Mgr. Irena Provazníková

Supervisor:

RNDr. Petr Nguyen, Ph.D.

Faculty of Science, University of South Bohemia in České Budějovice

České Budějovice 2022

This thesis should be cited as:

Provazníková, I. (2022) Drivers of karyotype evolution in Lepidoptera. Ph.D. Thesis. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 131 pp.

Annotation

Research of lepidopteran karyotypes and their evolution has been challenging for decades due to their many peculiar characteristics. However, this field has advanced thanks to modern cytogenetic techniques and sequencing technologies. We combined explored possibilities how to detect chromosomal rearrangements, and cytogenetic and genomic approaches to explore evolutionary forces shaping karyotypes of non-model Lepidoptera including representatives of early diverging species. Results obtained in the present thesis point to a possible role of satellite DNA and sexual antagonistic selection in mobilisation of rDNA and sex chromosome turnover, respectively.

Declaration

I hereby declare that I am the author of this thesis and that I have used only those sources and literature detailed in the list of references.

České Budějovice 24/11/2022

Irena Provazníková

This thesis originated from a partnership of Faculty of Science, University of South Bohemia, and Institute of Entomology, Biology Centre CAS, supporting doctoral studies in the Molecular and Cell Biology and Genetics study programme.

Financial support

This work was supported by:

Erasmus+ grant

The Czech Science Foundation, grant 17-17211S

The Czech Science Foundation, grant 20-20650Y

Acknowledgements

I would like to express my sincere gratitude to many people without whom this work would not be possible. First of all, I would like to thank my supervisor Peter for his support, wisdom, knowledge, motivation, and passion for science. I wish to extend my special thanks to all former and current members of Peter's and František's lab, especially Sandy, Marti, Anička, Lafi, Magda, and Leo for their endless support, and all the laughs and cries we have had over the years. Also, a big thanks goes to Maruška for taking such a good care of our moths and butterflies. Last but not least, the deepest gratitude belongs to my husband and my family for their love and support.

List of papers and author's contribution

The thesis is based on the following papers (listed chronologically):

I. Provazníková I., Hejníčková M., Visser S., Dalíková M., Carabajal-Paladino L.Z., Zrzavá M., Voleníková A., Marec F., Nguyen P. (2021) Large-scale comparative analysis of cytogenetic markers across Lepidoptera. Scietific Reports 11: 12214 doi: 10.1038/s41598-021-91665-7 (IF= 4.996)

IP designed experiments, performed research, processed and analysed data and wrote the first draft of the manuscript.

II. Dalíková M.*, Provazníková I.*, Provazník J., Grof-Tisza P., Pepi A., Nguyen P. (2022) The role of repetitive DNA in re-patterning of major rDNA clusters in Lepidoptera. bioRxiv 2022.03.26.485928; doi: 10.1101/2022.03.26.485928 (under revision in Genome Biology and Evolution)

IP designed experiments, performed FISH experiments and analyses of short reads by Repeat Explorer, analysed obtained data and wrote the first draft of the manuscript.

III. Carabajal Paladino L.Z.*, Provazníková I.*, Berger M., Bass C., Aratchige N. S., López S.N., Marec F., Nguyen P. (2019) Sex chromosome turnover in moths of the diverse superfamily Gelechioidea. Genome Biology and Evolution 11: 1307–1319. doi: 10.1093/gbe/evz075 (IF= 4.065)

IP analysed karyotypes of species under study by means of molecular cytogenetics, processed and interpreted obtained data and contributed to the first draft of the manuscript.

IV. Provazníková I., Dalíková M., Voleníková A., Roessingh P., Sahara K., Provazník J., Šlajsová M., Marec F., Nguyen P. (unpublished data) Multiple sex chromosomes of *Yponomeuta* ermine moths suggest a role of sexual antagonism in sex chromosome turnover in Lepidoptera.

IP performed FISH and qPCR experiments, analysed obtained data and contributed to the first draft of the manuscript.

* equal contribution

Co-author agreement

RNDr. Petr Nguyen, Ph.D., the supervisor of this Ph.D. thesis and corresponding author of papers I., II., III. fully acknowledges the stated contribution of Mgr. Irena Provazníková to these manuscripts and unpublished results section IV.

.....

RNDr. Petr Nguyen, Ph.D.

Contents

1. Introduction

1.1. General introduction

Speciation is a process by which populations evolve into different species. The earliest hypotheses regarding species formation were already formulated by philosophers in ancient Greece. However, for many centuries, occurrence of different plant and animal species was viewed as a result of the divine plan (Bardell, 1994; Li Causi, 2008). A major breakthrough occurred in 1858, when Charles Darwin and Alfred Wallace published a paper on the role of natural selection in evolution of species (Wallace and Darwin, 1858). Yet, the way traits are passed from parents to their offspring, and the origin of evolutionary novelties, remained a mystery for some time. Later on, German evolutionary biologist August Weismann described germ lines which give rise to egg and sperm cells as a mechanism of transmission of traits to offspring (Wallace, 1858). Mendel's laws of inheritance (Mendel, 1865) described the mechanistic basis of traits, however this work was not recognized at that time. In 1902 Walter Sutton and Theodor Boveri proposed the 'Chromosomal theory of heredity' explaining that chromosomes are the trait bearers. Later, Thomas Hunt Morgan published a book called *The Mechanism of Mendelian Heredity* (Morgan et al., 1915) which was based on his research on *Drosophila melanogaster* and combination of "Chromosomal theory" and Mendel's laws explaining the genetic basis of trait inheritance. Apart from the epigenetic inheritance which was discovered and defined later (Berger et al., 2009), all these and many more discoveries have contributed to the explanation that differences between all living organisms are encoded in their DNA (Gayon, 2016). In modern days we know that divergence on genomic level can lead to speciation, which is a source of extant biodiversity. Yet, despite being intensively studied, exact mechanisms and principles of this process are not fully understood due to its enormous complexity (Richards et al., 2019; Presgraves and Meiklejohn, 2021).

Nuclear DNA is usually comprised of several molecules called chromosomes, which are formed through protein-DNA interactions, and condense to become visible during cell division. A complete set of chromosomes from an individual organism constitutes a karyotype, which is characterized by number of chromosomes, their morphology, and constitution of sex chromosomes. Karyotypes can be specific and stable for certain species or populations (Wurster and Benirschke, 1970; Mank and Avise, 2006; Gonçalves-Oliveira et al., 2020) but inter- and intra- population variability is also common (Šíchová et al., 2015; Lorch et al., 2016). Number of chromosomes in eukaryotic genomes vary greatly, ranging from one chromosome per haploid genome (n=1) in the jack jumper ant *Myrmecia pilosula* (Crosland and Crozier, 1986) to ca. 720 chromosomes (n=720) found in the highly polyploid adder's tongue fern *Ophioglossum reticulatum* (Khandelwal, 1990).

Karyotypical changes caused by intra- and inter- chromosomal rearrangements (CRs) were proposed to play an important role in evolutionary processes (Leaché et al., 2016; Potter et al., 2017). The first CRs studied in detail were morphologically detectable fissions and fusions, for example fusion of two acrocentric chromosomes known as Robertsonian translocations, leading to a change in chromosome number. Also inversions can cause in visible changes in structure of chromosomes (Offner, 1994). Studying polytene chromosomes, i.e. large chromosomes containing a high number of fused sister chromatids, Dobzhansky (1950) observed heterozygous inversions of large chromosomal segments as loops in paired chromosomes in hybrid fruit flies (Dobzhansky, 1950).

However, structural changes are possible to detect only in species with prominent chromosomal landmarks, such as heterochromatin blocks (Chi et al., 2005) or a secondary constriction comprising tandem arrays of genes for major ribosomal RNAs (Henderson et al., 1972). Primary constrictions corresponding to centromeres, where spindle microtubules attach, are by far the most important feature, as its position is routinely used for morphological classification of monocentric chromosomes and can be compared between populations and species (Garagna et al., 2014; Robinson et al., 2014; Bracewell et al., 2019). CRs such as translocations or paracentric inversions, which do not comprise a centromere, may not cause morphologically visible changes and could remain undetected. Moreover, chromosomes of some organisms lack localized centromeres, and their spindle microtubules attach along chromosome surface, which hampers study of CRs. Nevertheless, both large scale and small scale CRs can be detected due to recent advances of molecular cytogenetic techniques and sequencing technologies, which allow comparative analysis in growing number of eukaryotic species (Deakin et al., 2019).

The work presented in this thesis is focused on changes in genome architecture of the order Lepidoptera, with emphasis on representatives of early diverging lineages which are generally understudied. In most cases, lepidopteran chromosomes are morphologically uniform and indistinguishable from one another, which makes their investigation very challenging (De Prins and Saitoh, 2003; Mediouni et al., 2004; Fuková et al., 2005). We used various means of fluorescence *in situ* hybridization (FISH) and bioinformatic tools to identify individual synteny blocks and study their evolution. Results of this work provide insights into karyotype and sex chromosome evolution within the order Lepidoptera.

1.2. Chromosomal speciation

The astonishing diversity of species we see on Earth today is a result of ongoing diversification (and extinction). To describe and explain the process, many models of various complexity have been developed. The complexity given by different factors such as mutation, genetic drift, sexual

selection, non-random mating and recombination influences the evolutionary dynamics in various ways (Gavrilets, 2014).

Build-up of reproductive barrier between two geographically isolated populations, i.e. in allopatry, can be explained by the Bateson-Dobzhansky-Muller (BDM) model of genetic incompatibilities. The widely recognized BDM model describes how two populations genetically diverge from each other upon geographic separation. Their genomes evolve independently and become genetically incompatible, which can reduce hybrid fitness when secondary contact of the populations occurs (reviewed in Gavrilets, 2014). The general BDM model has been further developed (e.g. Barton and de Cara, 2009; Fierst and Hansen, 2010; Fraïsse et al., 2017) and speciation in the absence of gene flow has been considered a feasible mechanism (Bolnick and Fitzpatrick, 2007). In theory, CRs accumulated in allopatry could additionally induce hybrid incompatibilities as they could reduce fitness of hybrid offspring (heterokaryotypes) via incorrect meiotic segregation and production of unbalanced gametes (Faria and Navarro, 2010) . However, these theoretical models gained little empirical support and were challenged, as CRs causing severe fitness reduction would likely never become fixed in diverging populations in the first place (Rieseberg, 2001; Faria and Navarro, 2010).

Formation of reproductive isolation with gene flow between two populations with the same or overlapping geographic distribution, i.e. sympatric and parapatric speciation, respectively, has been more difficult to explain (Fitzpatrick et al., 2008; Richards et al., 2019). New theoretical models have revived interest in CRs as they assume that BDM incompatibilities can accumulate in the face of gene flow in regions with suppressed recombination. The rearranged chromosomal regions experience reduced recombination in heterozygotes and therefore can shelter allelic associations underlying reproductive isolation or local adaptation (Noor et al., 2001; Rieseberg, 2001; Navarro and Barton, 2003; Kirkpatrick and Barton, 2006; Faria and Navarro, 2010).

In eukaryotes, adaptation can be facilitated by a wide range of genomic differences, from insertion of mobile element (ME) into a single gene (van't Hof et al., 2016), through large chromosomal rearrangements (Sember et al., 2015), to whole genome duplication (Baduel et al., 2018). One of the many types of CRs contributing to ecological adaptation are gene duplications. Since the original gene already performs a certain function, its redundant duplicate can gain a completely new function or enhance the effect of the original function (Ohno, 1970; Kondrashov, 2012). For example, the acquired ability of aphid *Myzus persicae* was found to be facilitated by several small- and large- scale mutation events which caused significant amplification of genes encoding detoxifying enzymes causing resistance to nicotine, the tobacco's defense against pests.

Interestingly, higher expression of these genes was found in aphid gut as well as bacteriocyte containing endosymbiotic bacteria. This underlines the complexity of the ecological adaptation in phytophagous insects, which, in this case, led to ecological diversification as *Myzus spp. nicotianae* is recognized as a distinct subspecies of *M. persicea* (Singh et al., 2020).

Chromosomal inversions represent one of the most significant contributors to ecological adaptation in the presence of gene flow (Twyford and Friedman, 2015). Species can have a single (Fang et al., 2012) or multiple inversions spread throughout the genome (Morales et al., 2019). Chromosomal inversion is responsible for recombination suppression in heterokaryotypes. The lack of genetic exchange between these regions then results in linkage disequilibrium (LD) among loci within the inverted region, which can maintain association of alleles which provide advantage in various environmental conditions and drive genetic divergence (reviewed in Hoffmann and Rieseberg, 2008). There are two classic examples of the role of inversions in ecological adaptation. Lowry and Willis, (2010) discovered an inversion polymorphism fixed in two ecotypes of the yellow monkeyflower *Mimulus guttatus,* contributing to local adaptation. The alleles in the inverted region influence morphology and flowering time and are linked to the different biotopes. The second example is the mosquito *Anopheles gambiae*, the notorious malaria vector in Africa. In several populations of this species, different paracentric chromosomal inversions were observed. The inverted chromosome piece contains alleles responsible for adaptation to aridity, feeding preference and insecticide resistance to DDT and dieldrin (Brooke et al., 2002). Thanks to advances in sequencing technologies, such inversions have been recently found and described in detail in a wide range of species (Fang et al., 2012; Christmas et al., 2019; Mérot et al., 2021; Matschiner et al., 2022).

Chromosomal fusions can bring together previously unlinked adaptive loci creating novel combinations (Guerrero and Kirkpatrick, 2014), and facilitate the adaptation process by significantly changing the recombination rate and background selection in fused chromosomes (Cicconardi et al., 2021). The most common types of chromosomal fusion in animals and plants are Robertsonian translocations which can alter not only the recombination landscape but also spatial genome folding, thus influencing gene expression (Vara et al., 2021). Even though chromosomal fusions are less studied than chromosomal inversions, there is empirical evidence of their important role in adaptation (Wellband et al., 2019; Cicconardi et al., 2021; Liu et al., 2022). For example, Liu et al., (2022) found two autosomal fusion events in the threespine stickleback genome. The fused chromosomes display low recombination rate, hypothetically protecting the newly formed clusters of adaptive loci in the face of recurrent gene flow between freshwater and marine populations. Besides autosomal fusions, fusions involving sex chromosomes (S-A fusions) seem to be relatively

common and were shown to play an important role in evolutionary processes such as adaptation (Nguyen et al., 2013), sexual antagonism and speciation (Smith et al., 2016) by bringing previously autosomal loci under beneficial sex linkage.

1.3. Sex chromosomes

Sex chromosomes have originated independently and numerous times at various evolutionary rates during the evolution of Eukaryotes (Mrackova et al., 2008; Takehana et al., 2008; Veyrunes et al., 2008; Blackmon et al., 2017). They are divided into two main systems. Male heterogamety, with the XY/XX system (and its variants), is typical for most insects, mammals, some fish, reptiles, amphibians, and nematodes. Conversely, female heterogamety, with the ZW/ZZ system (and its variants), is found exclusively only in birds and the two sister insect orders of Lepidoptera and Trichoptera, whereas there are only anecdotic cases of female heterogamety in the other taxa (Bachtrog et al., 2014; Blackmon et al., 2017 and references within).

According to the classical model, sex chromosomes evolve from a pair of autosomes as one of the homologues acquires a master sex-determining locus (Charlesworth, 1978). Next, sexually antagonistic alleles (beneficial to one sex but harmful to the other) accumulate adjacent to the sex-determining locus. During these early stages, a recombination between the regions of the nascent sex chromosomes is expected to be suppressed due to sexual conflict. Sexually antagonistic alleles beneficial to males, for example, accumulate near the male-determining locus on the nascent Y chromosome and become the primary drivers of recombination suppression between the sex chromosomes in males. In this way, the non-recombining Y chromosome becomes male-specific, while recombination of X chromosomes in females occurs normally without any suppression (reviewed in Ellegren, 2011; Furman et al., 2020). However, according to an alternative hypothesis, sex-specific and antagonistic genes are translocated to sex chromosomes, or develop sex-specific function after recombination is restricted (reviewed in Wright et al., 2016). Some recent models predict that loss of recombination and sequence divergence are caused by accumulation of neutral differences without the sexual antagonistic selection (Jeffries et al., 2021; Jay et al., 2022). The sex determining region, and level of divergence between X and Y or Z and W, can be very small. For example, in tiger pufferfish, only a single missense SNP (single nucleotide polymorphism) on the proto-Y is associated with male development (Kamiya et al., 2012). It can also expand, and recombination then becomes restricted along almost the entire chromosome (Charlesworth et al., 2005). The absence of recombination results in a decay of sex-limited chromosomes via so-called Hill-Robertson effects, such as an accumulation of weakly deleterious mutations through Muller's ratchet. The accumulation of such mutations and repetitive sequences contribute to

pseudogenization of genes and heterochromatinization of the sex-limited sex chromosomes Y or W, respectively (Charlesworth and Charlesworth, 2000; Kaiser and Bachtrog, 2010; Bachtrog, 2013).

Alternatively, the established sex chromosome systems can be also converted from XY to ZW and vice versa, as the new sex determining locus takes over the role of the sex-determining pathway (Meisel, 2020). The W and Y chromosomes were also found to arise from a supernumerary B chromosome in cichlid fish (Clark and Kocher, 2019) and blind cavefish (Imarazene et al., 2021), respectively, or by incorporating a feminizer from Wolbachia into the genome of the common pillbug *Armadillium vulgare* (Leclercq et al., 2016).

Sex chromosomes are also subject to frequent CRs such as chromosomal fusions resulting in so-called neo-sex chromosomes. One of the most bizarre examples of multiple sex chromosome constitution is found in a representative of monotremes, the duck-billed platypus, *Ornithorhynchus anatinus*. As all therian mammals, platypus also have a sex chromosome system with male heterogamety. However, it consists of five X chromosomes and five Y chromosomes in males, and five pairs of X chromosomes in females, which is the highest number of sex chromosomes found in mammals (Rens et al., 2004) originating from an ancestral chromosome ring formation (Zhou et al., 2021). Unlike the stable and old heteromorphic sex chromosomes of mammals (Bachtrog, 2013; Cortez et al., 2014), a rapid turnover of homomorphic sex chromosomes was found in teleost fish, especially in African cichlids (Kitano and Peichel, 2012; Gammerdinger and Kocher, 2018). Sex chromosome systems X0/XX, XY/XX, Z0/ZZ, ZW/ZZ and their variants and polygenic systems, or feminizing role of B chromosomes, were observed in fishes (reviewed in Gammerdinger and Kocher, 2018). Their extraordinary number of sex chromosome systems and sex-determination pathways is often linked with their sexual dimorphism (Roberts et al., 2009) and their turnover was suggested to even promote speciation (Kitano et al., 2009; Kitano and Peichel, 2012).

1.4. Lepidopteran karyotype

Lepidopteran mitotic chromosomes are morphologically uniform, which makes it difficult to differentiate individual elements (Mediouni et al., 2004; Fuková et al., 2005). Moreover, classical cytological techniques failed to provide any chromosome-specific banding patterns (Bedo, 1984; De Prins and Saitoh, 2003). Therefore, research was limited only to counting of mitotic chromosomes using orcein or Giemsa staining techniques (Robinson, 1971). Although using just simple staining techniques, karyotypes have been described in hundreds of lepidopteran species providing valuable insight into general trends of their karyotype evolution (Robinson, 1971; Lukhtanov, 2000). A majority of moths and butterflies have a haploid chromosome number close to n=30 (Robinson, 1971; Ahola et al., 2014) and the ancestral chromosome number was determined to be n=31 by

comparative genomics studies (Van't Hof et al., 2013; Ahola et al., 2014). However, Lepidoptera have the highest variation in chromosome number of all diploid animals (Hill et al., 2019; Mandrioli and Manicardi, 2020) as evidenced by species with extreme chromosome numbers. On one hand, the Arizona giant skipper *Agathymus aryxna* (Hesperiidae) and *Hypothyris thea* (Nymphalidae), both with haploid chromosome number n=5, represent lepidopterans with the lowest chromosome number (De Prins and Saitoh, 2003; Brown Jr et al., 2005). On the other hand, the Atlas blue butterfly *Polyommatus atlantica* (Lycaneidae) is, with n=224-226, the species with the highest non-polyploid chromosome number in the animal kingdom (Lukhtanov, 2015). Moreover, the genus *Polyommatus* is known to contain three independent clades with extremely high chromosome numbers caused by chromosome fission events (reviewed in Kandul et al., 2007; Lukhtanov et al., 2020). It seems that lepidopterans can cope with genome fragmentation and chromosome fusions relatively well. It has been proposed that changes in lepidopteran chromosomal number could be facilitated by their holocentric chromosomes (Wolf et al., 1997). The holocentric chromosomes present in phytophagous insects, such as some aphids, moths and butterflies, could have arisen as adaptation against clastogenic compounds from host plants preventing damage to their DNA (Mandrioli and Manicardi, 2020). De Vos et al. (2020) proposed that changes in chromosomal numbers in Lepidoptera are not associated with speciation. However, detailed studies in genera *Polyomatus* and *Leptidea* showed that karyotype differences between closely related species could facilitate speciation in sympatry, and that high chromosome numbers correlate with high diversification rates (Kandul et al., 2007; Talavera et al., 2013; Lukhtanov, 2015).

Since lepidopteran chromosomes are basically indistinguishable from one another by conventional staining techniques (Bedo, 1984; De Prins and Saitoh, 2003; Fuková et al., 2005), FISH became a very powerful method, which overcome these issues and enabled the study of lepidopteran karyotypes in detail (Vítková et al., 2007; Nguyen et al., 2010; Yoshido et al., 2020). In general, FISH and its modifications are widely used for comparative evolutionary studies in plants (e.g. Široký et al., 2001; Macas et al., 2007; Lusinska et al., 2018) and animals (e.g. Rens et al., 2004; Azevedo et al., 2012; Cavalcante et al., 2018), and proved to be useful for identification of cryptic species (Šíchová et al., 2015), and hybrids in nature (Lukhtanov, 2015). This relatively simple and cost-effective technique allows mapping and visualization of a probe derived from a sequence of interest on the target chromosomes. Fluorescently labelled probes can be prepared against whole genomes (Mongue et al., 2017), specific chromosomes (painting probes) (Hejníčková et al., 2021), various repetitive sequences (Cabral-de-Mello et al., 2021), gene families (Nguyen et al., 2010; Šíchová et al., 2013) or even single copy genes (Carabajal Paladino et al., 2014).

The so-called universal cytogenetic markers, such as telomeres or rRNA genes, have conserved sequences and thus can be successfully hybridized to closely related species, representatives of different superfamilies, or even different orders which allows the extension of cytogenetic studies to numerous non-model organisms (Vítková et al., 2005; Macas et al., 2007; Nguyen et al., 2010; Aguilera et al., 2016). For some of the markers, there is no need for prior knowledge of genome sequences of studied taxa. In Lepidoptera, a fragment of the 18S rDNA gene obtained from the codling moth *Cydia pomonella* (Tortricidae) (Fuková et al., 2005) was successfully mapped in many other species across the whole order (Nguyen et al., 2010; Šíchová et al., 2013, 2015). Also, FISH with telomeric repeats has been used to highlight chromosome ends, which facilitates chromosome counting and identification of multivalents, e.g. in the *Leptidea* species (Šíchová et al., 2015, 2016).

Physical chromosome maps reveal distribution of selected markers, i.e. size, number, and position of hybridization signals, and can be compared between taxa or used to infer evolution of specific parts of the genome (Nguyen et al., 2010; Cabral-de-Mello et al., 2011; Rovatsos et al., 2019; Yoshido et al., 2020). Using FISH with bacterial artificial chromosomes (BACs) or fosmids as probes, physical maps of genomes were obtained and used in comparative analyses of synteny of genes, and in chromosome barcoding in plants (Lusinska et al., 2018) and animals (Mendoza et al., 2020) including lepidopterans (e.g Yasukochi et al., 2009; Yoshido et al., 2011; 2020). BACs and fosmids contain large inserts (tens to hundreds of kbp) of genomic DNA (gDNA) of species of interest and can be detected by basic FISH (BAC-FISH, fosmid-FISH) (Yasukochi et al., 2009; Yoshido et al., 2011). In the case of Lepidoptera, Yoshido et al., (2005a) used BAC-FISH to identify all 28 chromosomes of the model species, domestic silk moth *Bombyx mori* (Bombicidae). Later on, this map was compared to a physical map of the genome of tobacco hornworm *Manduca sexta* (Sphingidae). Comparative analyses of physical maps revealed few CRs, but also highly conserved synteny, including gene order, between genomes of *B. mori* and *M. sexta* (Yasukochi et al., 2009). Similar results showing a high level of synteny were also observed between BAC- and fosmid- based physical maps of *B. mori* and the European corn borer *Ostrinia nubilalis* (Crambidae, n= 31) also proving n=31 to be the ancestral haploid chromosomal number (Yasukochi et al., 2016). Recently, Yoshido et al. (2020) used BAC-FISH to identify all chromosomes involved in sex chromosome multivalents in three *Leptidea* species known for highly unstable genomes (Šíchová et al., 2015). As these techniques are relatively time-consuming and costly, especially for species with large genomes, such maps have been constructed for just a handful of species and have been replaced by chromosome level genome assemblies.

In 2004, the first lepidopteran genome of *B. mori* was sequenced by Sanger technology and assembled (Mita et al., 2004; Xia et al., 2004). During the next few years, the silkworm genome was

further improved by an integration of physical map and linkage maps based on expressed sequence tags (EST) (Yasukochi et al., 2008) and SNPs (Yamamoto et al., 2008). A resulting high-quality assembly and linkage map of *B. mori* served as a reference for comparative analyses focused on chromosomal evolution in several non-model lepidopteran species (e.g. Van't Hof et al., 2013; Ahola et al., 2014; Hill et al., 2019). However, a major revolution in the field of comparative genomics has been brought about by the advent of next generation sequencing (NGS) such as the second generation Illumina sequencing and third generation (single molecule long-read) sequencing technologies of Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio), allowing large-scale comparative analyses of tens of specimens, and identification of large inter- and intra-CRs (Conte et al., 2019; Kautt et al., 2020). Having a high-quality reference genome is an essential part of such comparative studies (Conte et al., 2019). Due to the decreasing price of sequencing, reference genomes are no longer available only for a handful of models but for many non-model species as well (Ma et al., 2018; Hill et al., 2019). Comparative genome analyses were carried out across ditrysian species from superfamilies Papilionoidea (Beldade et al., 2009; Ahola et al., 2014), Geometroidea (Van't Hof et al., 2013), and Tortricoidea (Picq et al., 2018; Wan et al., 2019). Despite various chromosome numbers, synteny blocks corresponding to chromosomes in an ancestral karyotype seem to be highly conserved. The only exception so far is the green-veined white *Pieris napi* (Papilionoidea) whose genome is extensively rearranged due to multiple CRs (Hill et al., 2019).

1.5. Lepidopteran sex chromosomes

As mentioned above, moths and butterflies represent one of few lineages with exclusive female heterogamety. In fact, Lepidoptera represents, with their ca. 160,000 described species (Van Nieukerken et al., 2011), the largest group with female heterogamety in the animal kingdom. The ZW/ZZ sex determination system is otherwise relatively rare in insects as it evolved independently only in the fruit fly family Tephritidae and in a common ancestor of the superorder Amphiesmenoptera comprising Lepidoptera and Trichoptera (Blackmon et al., 2017). The predominant constitution of sex chromosomes in Lepidoptera is ♀ZW/♂ZZ (Sahara et al., 2012). However, ♀Z0/♂ZZ is considered to be the ancestral state, as it was observed mainly in early diverging lepidopteran lineages and in the sister order Trichoptera (Marec and Novak, 1998; Lukhtanov, 2000; Traut et al., 2007). It was hypothesised that the W chromosome evolved in a common ancestor of the family Tischeriidae and the group Ditrysia (containing 98% of lepidopteran species). Two main scenarios were proposed to explain evolutionary origin of the W chromosome. According to the first, the W chromosome evolved from an autosome, the homologue of which fused with the ancestral Z chromosome (Traut and Marec, 1996). As the newly arisen W chromosome became sex-limited and recombination with Z chromosome was supressed due to

female achiasmatic meiosis, it started accumulating repetitive sequences and degenerating, in agreement with the canonical model (Charlesworth, 1978). The second scenario explains the origin of the W chromosome via an acquisition of a B chromosome which started to pair with the Z chromosome (Lukhtanov, 2000). Presence or absence of the W chromosome was determined in many species only indirectly by detecting so-called sex chromatin or W chromatin. Sex chromatin is a stainable heterochromatic body, which consists of multiple copies of the W chromosome in interphase nuclei of polyploid cells (Traut and Marec, 1996). However, some studies indicate that sex chromatin is not a reliable proxy for presence of the W chromosome (Hejníčková et al., 2019; Voleníková, 2015). Moreover, absence of the W chromosome in representatives of early-diverging ditrysian lineages suggests several independent origins of the W chromosome during lepidopteran evolution (Dalíková et al., 2017a). Interestingly, a secondary loss of the W chromosome was detected in advanced lepidopteran families, in *Samia cynthia ricini* (Saturnidae, Yoshido et al., 2005) and in *Aethalura punctulata* (Geometridae, Hejníčková et al., 2021). The lepidopteran W chromosomes are usually highly heterochromatic and comprise mostly repetitive sequences, particularly various MEs (Fuková et al., 2007; Wan et al., 2019). Only a few protein-coding genes have been localized to the W chromosome so far (Gotter et al., 1999; Nagaraju et al., 2014). Interestingly, sex determining *Feminizer* piRNA was found on the W chromosome of *B. mori* (Kiuchi et al., 2014).

Comparative genomic hybridization (CGH) has been widely used in cytogenetic research of the lepidopteran W chromosome by hybridizing differently labelled female and male whole-genome probes on female chromosome preparations (Mediouni et al., 2004; Fuková et al., 2005; Vítková et al., 2007; Šíchová et al., 2013; Mongue et al., 2017; Hejníčková et al., 2021). In addition to CGH, a variety of FISH methods with different types of probes such as BAC-FISH (Yoshido et al., 2005a), female-derived probes (genomic *in situ* hybridization, GISH) (Sahara et al., 2003; Carabajal Paladino et al., 2019), or W-painting probes obtained by laser microdissection of sex chromatin bodies (Fuková et al., 2007) have been used to highlight the W chromosome. Moreover, a combined use of GISH with telomeric probes revealed the presence of multiple sex chromosome systems in several ditrysian species (Sahara et al., 2003; Šíchová et al., 2015). However, the composition of repetitive sequences and their distribution on the W chromosome differ among closely related species, which indicates dynamic evolution of the W chromosome (Yoshido et al., 2005b; Vítková et al., 2007). This hampers comparative analysis and investigation of the W origin across Lepidoptera (Dalíková et al., 2017a). High abundance of repetitive content and differences in molecular composition of the W chromosome among species make their comparative analyses very challenging, not only for cytogenetic but also for bioinformatic analyses (Wilson Sayres, 2018; Lewis et al., 2021). The W

chromosome was therefore overlooked or even excluded from many sequencing projects. However, the long reads generated by third generation sequencing technologies, combined with other approaches such as Hi-C, have great potential to improve assemblies of W chromosomes, provide additional information about their structure, and shed light on their origins and evolution (Wan et al., 2019; Lewis et al., 2021). Nevertheless, a high-quality assembly of W chromosomes is yet to come.

In contrast to the W chromosome, the Z chromosome has been well studied due to its autosome-like features, such as prominent euchromatin presence and rich gene content (Van't Hof et al., 2013). The Z chromosome evolved in a common ancestor of Lepidoptera and Trichoptera. It is conserved in early diverging Lepidoptera (Dalíková et al., 2017a; Fraïsse et al., 2017), as well as in advanced Ditrysia (Yasukochi et al., 2009; Van't Hof et al., 2013; You et al., 2013; Ahola et al., 2014). During the last decade, it has been shown that fusions of Z chromosomes with autosomes resulting in neo-Z chromosomes are relatively common within Lepidoptera (Nguyen et al., 2013; Dalíková et al., 2017a; Fraïsse et al., 2017; Mongue et al., 2017; Picq et al., 2018; Hill et al., 2019). Interestingly, Pennell et al. (2015) showed that in fish and reptiles, where both XX/XY and ZW/ZZ systems are found, Y-autosome fusions have the highest establishment rate compared to W-autosome, X-autosome, and Z-autosome fusions. It seems that Z-autosome or W-autosome fusions are generally rare in vertebrates (The Tree of Sex Consortium, 2014; Pokorná et al., 2014; Pennell et al., 2015). However, the increasing number of species with neo-sex chromosomes could extend this relatively short list and shed some light on the evolutionary forces driving fusions (e.g. Mongue et al., 2017; Huang et al., 2022). Hypothetically, Z-linkage of previously autosomal genes provides a great benefit due to the "Fast-Z effect", whereby recessive mutations are exposed to positive and purifying selection on the Z chromosome in hemizygous sex, which then undergoes faster molecular evolution compared to autosomes (Charlesworth et al., 2005). Sex chromosomes are thus a great hub for adaptive genes. For instance, in leafrollers of the family Tortricidae, Z chromosome fused with an autosome corresponding to chromosome 15 in the *B. mori* genome. (Nguyen et al., 2013). It was hypothesized that the Z chromosome -autosome fusion in tortricids was fixed due to advantageous linkage between autosomal larval performance and Z-linked female oviposition preference or reproductive isolation, which may have contributed to their ecological diversification and consequent radiation (Nguyen et al., 2013; Picq et al., 2018).

In addition to the basic constitution and neo-sex chromosomes, multiple sex chromosomes have also been detected in various lepidopteran groups (Nilsson et al., 1988; Šíchová et al., 2015, 2016; Hejníčková et al., 2021). For example, four species of the genus *Leptidea* have one of the most complex sex chromosome systems described in Lepidoptera: $W_{1–3}Z_{1–4}$/ $Z_{1–4}Z_{1–4}$ in *L. juvernica*, $W_{1–3}Z_{1–}$

$_3$/$Z_{1–3}Z_{1–3}$ in *L. sinapis*, $W_{1–4}Z_{1–4}$/$Z_{1–4}Z_{1–4}$ in *L. reali* and $W_{1–3}Z_{1–6}$ /$Z_{1–6}Z_{1–6}$ in *L. amurensis* (Šíchová et al., 2015, 2016). Recently Yoshido et al., (2020) identified synteny blocks corresponding to *B. mori* chromosome Z and autosomes 7, 8, 11, 15, 17 and 24 to be involved in sex chromosome multivalents of *L. juvernica*, *L. sinapis* and *L. reali*. New Z-linked genes could play an important role in divergence and speciation via an accumulation of genetic incompatibilities between populations (Šíchová et al., 2015; Yoshido et al., 2020).

2. Outline of research

The species-rich order Lepidoptera (moths and butterflies), containing numerous economically important pest species, represents a great model for physiological, ecological, evolutionary, genetic and developmental studies (Roe et al., 2009). Lepidopteran karyotypes have been relatively difficult to study due to their peculiarities such as dot-shaped chromosomes, a lack of localized centromere, and a failure of banding techniques (De Prins and Saitoh, 2003). However, thanks to the ongoing progress of cytogenetic techniques and sequencing technologies it is possible to examine genome and karyotype evolution in many non-model species and identify various CRs and forces driving karyotype evolution of Lepidoptera (Nguyen et al., 2013; Van't Hof et al., 2013; Hill et al., 2019; Yoshido et al., 2020). The main goal of this thesis is to advance our understanding of karyotype evolution in Lepidoptera, which is a much needed prerequisite of our understanding of a role of CRs in evolution.

Universal cytogenetic markers have been widely used in comparative evolutionary studies, and for chromosome barcoding of non-model species. The most used and established marker is the 18S rDNA gene of the major rDNA (e.g. Nguyen et al., 2010; Cabral-de-Mello et al., 2011). Its application is robust due to a organization of hundreds of copies in tandem arrays and conserved nucleotide sequence (Prokopowich et al., 2003; Kobayashi, 2006) due to which it can be used in many non-model species without any prior knowledge of genome of species under study. Histone genes, 5S rDNA or U snRNA genes have been also used as cytogenetic markers, however, not to the same extent as the 18S rDNA gene. The use of these markers separately or in a combination has a great potential in studies of karyotype evolution (Cabral-de-Mello et al., 2010). In the first chapter, we analysed karyotype changes in selected species using two well-established cytogenetic markers in lepidopteran research, namely 18S rDNA and histone H3 (e.g. (Fuková et al., 2005; Nguyen et al., 2010; Šíchová et al., 2015), and three markers, namely 5S rDNA, U1 and U2 snRNA genes, which has never been mapped in Lepidoptera before. Besides analysing the distribution of the markers in genomes using FISH and Southern blot, we also estimated their copy numbers by qPCR and evaluated their applicability in lepidopteran comparative cytogenetics. The histone H3 mapping revealed relatively conserved cluster organization in majority of studied species with only few changes in number and position of the H3 loci between species. Besides few exceptions, the analyses of 5S rDNA, U1 and U2 snRNA genes revealed their scattered organization and low copy numbers in lepidopteran genomes and thus these markers are not suitable to study karyotype evolution in moths and butterflies. On the contrary, hybridization pattern of 18S rDNA genes showed dynamic evolution of the major rDNA cluster throughout the order, with one terminal cluster per diploid genome as an ancestral state.

Even though a single major rDNA cluster per diploid genome is the most common in animals, high numbers of clusters and unusual derived distribution patterns were also observed (Sochorová et al., 2018). Therefore, in the second chapter we focused on extraordinary patterns of major rDNA distribution detected in several lepidopteran species, namely *Hepialus humuli* and *Cameraria ohridella* with one extremely large rDNA cluster and nymphalids *Aglais urticae* and *Inachis io* with multiple small clusters. The question at hand was how the rDNA arrays get mobilized and ultimately, what factors could facilitate CRs in Lepidoptera. Interestingly, various satellites and ME were found inserted within the intergenic spacer (IGS) or the rDNA genes. The results point to a role of satellite DNA in repatterning of the major rDNA in Lepidoptera and highlight the advantages of combined cytogenetic techniques and sequencing technologies in studies of repetitive sequences.

The order Lepidoptera together with its sister order Trichoptera represents the largest groups with female heterogamety. Lepidopteran sex chromosomes are in general easy to identify by means of cytogenetics, either due to differences in their morphology or differences in chromosome numbers between sexes in case of either Z0 or multiple sex chromosome systems (Hejníčková et al., 2021).

In the third chapter, we analyzed sex chromosomes of pests of the speciose superfamily Gelechioidea. Analysis of karyotypes of several gelechiid representatives revealed reduced chromosome numbers, most probably caused by autosome fusions as one or more large chromosome pairs were observed (Ennis, 1976). However, an analysis of the karyotype of the tomato leaf miner *Tuta absoluta,* identified the largest chromosome pair as sex chromosomes (Carabajal Paladino et al., 2016). We examined in detail sex chromosomes of *T. absoluta* and other representatives sampled across the Gelechioidea superfamily. We showed that the sex chromosomes of *T. absoluta* contain synteny blocks corresponding to the autosome 7 (LG7) and 27 (LG27) of *Biston betularia* (Geometridae) representing the ancestral genome architecture with n=31 chromosomes (Van't Hof et al., 2013). Further, we showed that the fusion between the sex chromosomes and the autosome LG7 occurred in a common ancestor of the whole superfamily, whereas the second fusion occurred later and is specific only to the tribe Gnoreschemini. Cytogenetic data further suggest yet another S-A fusion in the family Oecophoridae. We argued that the gelechioid neo-sex chromosomes well evidence high incidence of neo-sex chromosomes in Lepidoptera. Hence, the paucity of S-A fusions is not an intrinsic feature of female heterogamety as previously proposed for vertebrates (Pokorná et al., 2014; Pennell et al., 2015).

In the fourth chapter, we studied multiple sex chromosome system of the small ermine moths of the genus *Yponomeuta* (Yponomeutoidea). Nilsson et al. (1988) analyzed karyotypes of

several ermine moth species and revealed difference in chromosome number between males (2n=62) and females (2n=61) and a presence of a trivalent in female oocytes formed by two Z chromosomes and one large W chromosome (♀Z_1Z_2neo-W/ ♂$Z_1Z_1Z_2Z_2$). We performed karyotype analyses and confirmed the presence of the trivalent in 9 out of 10 studied *Yponomeuta* species sampled across the whole genus. Interestingly, in case of *Y. tokyonella*, the Z_1 and the Z_2 chromosomes further fused forming the neo-Z chromosome pairing with the neo-W chromosome. Moreover, we found out that the Z_2 chromosome of *Y. evonymella* corresponds to the chromosome 2 of *B. mori* (BmCh2), and proved that the multiple sex chromosome constitution occurred in a common ancestor of the genus *Yponomeuta* and the outgroup, *Teinoptila gutella*. The BmChr2 bears major cluster of genes with ovary specific expression (Suetsugu et al., 2013). As the homeologs of other *B. mori* chromosomes enriched in ovary specific genes fused with sex chromosomes in moths of the family Tortricidae (Nguyen et al. 2013) and *Danaus* butterflies (Mongue et al. 2017), we hypothesize that sexual antagonism is the driving force behind the sex chromosome turnover in Lepidoptera.

3. Original publications

3.1. Chapter I

Large-scale comparative analysis of cytogenetic markers across Lepidoptera.

Provazníková I., Hejníčková M., Visser S., Dalíková M., Carabajal-Paladino L.Z., Zrzavá M., Voleníková A., Marec F., Nguyen P.

Scietific Reports 11: 12214 (2021), doi: 10.1038/s41598-021-91665-7

Abstract

Fluorescence in situ hybridization (FISH) allows identification of particular chromosomes and their rearrangements. Using FISH with signal enhancement via antibody amplification and enzymatically catalysed reporter deposition, we evaluated applicability of universal cytogenetic markers, namely 18S and 5S rDNA genes, U1 and U2 snRNA genes, and histone H3 genes, in the study of the karyotype evolution in moths and butterflies. Major rDNA underwent rather erratic evolution, which does not always reflect chromosomal changes. In contrast, the hybridization pattern of histone H3 genes was well conserved, reflecting the stable organisation of lepidopteran genomes. Unlike 5S rDNA and U1 and U2 snRNA genes which we failed to detect, except for 5S rDNA in a few representatives of early diverging lepidopteran lineages. To explain the negative FISH results, we used quantitative PCR and Southern hybridization to estimate the copy number and organization of the studied genes in selected species. The results suggested that their detection was hampered by long spacers between the genes and/or their scattered distribution. Our results question homology of 5S rDNA and U1 and U2 snRNA loci in comparative studies. We recommend the use of histone H3 in studies of karyotype evolution.

scientific reports

Check for updates

OPEN

Large-scale comparative analysis of cytogenetic markers across Lepidoptera

Irena Provazníková[1,2,3], Martina Hejníčková[1,2], Sander Visser[1,2,4], Martina Dalíková[1,2], Leonela Z. Carabajal Paladino[5], Magda Zrzavá[1,2], Anna Voleníková[1,2], František Marec[2] & Petr Nguyen[1,2]✉

Fluorescence in situ hybridization (FISH) allows identification of particular chromosomes and their rearrangements. Using FISH with signal enhancement via antibody amplification and enzymatically catalysed reporter deposition, we evaluated applicability of universal cytogenetic markers, namely 18S and 5S rDNA genes, U1 and U2 snRNA genes, and histone H3 genes, in the study of the karyotype evolution in moths and butterflies. Major rDNA underwent rather erratic evolution, which does not always reflect chromosomal changes. In contrast, the hybridization pattern of histone H3 genes was well conserved, reflecting the stable organisation of lepidopteran genomes. Unlike 5S rDNA and U1 and U2 snRNA genes which we failed to detect, except for 5S rDNA in a few representatives of early diverging lepidopteran lineages. To explain the negative FISH results, we used quantitative PCR and Southern hybridization to estimate the copy number and organization of the studied genes in selected species. The results suggested that their detection was hampered by long spacers between the genes and/or their scattered distribution. Our results question homology of 5S rDNA and U1 and U2 snRNA loci in comparative studies. We recommend the use of histone H3 in studies of karyotype evolution.

Cytogenetic studies aim at characterization of genome organization and its changes. Previously indispensable for the identification of genes of interest, cytogenetics may seem to struggle in the post-genomic era as it lags behind the resolution of molecular biology and genomics. Yet it remains crucial for genomic research. Cytogenetic data such as genome size and chromosome number allow for an informed choice of sequencing strategies and provide hypothetical framework for genomic studies, context to bioinformatic analyses, and physical evidence for results produced in silico[1–3]. Recent efforts, such as the Earth BioGenome project that aspire to characterize genomes of all eukaryotic biodiversity[4], will without a doubt lead to further cytogenetic research. As a result, the new field integrating cytogenetics and genomics has recently been proposed under the term chromosomics (coined by Claussen[5] but repurposed later by Graphodatsky[6] and Deakin et al.[3]).

There are several approaches to distinguish individual chromosomes within a karyotype. Classical techniques such as orcein or Giemsa staining as well as various banding methods can produce chromosome-specific patterns. These techniques work very well in mammals including humans[7,8], other vertebrates[9], some invertebrate taxa[10–12] and plants[13,14]. However, classical staining and banding techniques have failed in some organisms, such as moths and butterflies[15,16].

Lepidoptera with more than 160,000 described species and great ecological diversity[17] represent an excellent model system to study karyotype evolution and the role of changes in genome architecture in evolutionary processes. In Lepidoptera, chromosomal rearrangements such as inversions, fusions, and fissions play an important role in speciation[18] and adaptation, such as resistance to insecticides[19,20] and baculoviruses[21] and detoxification of plant secondary metabolites[22] and xenobiotics[23]. However, comparative cytogenetic studies are scarce in Lepidoptera due to the many peculiarities of lepidopteran chromosomes. Mitotic complements of both Lepidoptera and their sister group Trichoptera typically consist of a high number of small and morphologically uniform holokinetic chromosomes[24,25]. Since they lack a primary constriction, i.e. the centromere, its position cannot be used in chromosome identification[26]. Thus, cytogenetic analyses of lepidopteran karyotypes were challenging for years before molecular cytogenetic tools were introduced[25,27] and applied on meiotic pachytene chromosomes rather than mitotic chromosomes[28]. However, broader comparative cytogenetic studies, which would help us to

[1]Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. [2]Institute of Entomology, Biology Centre CAS, České Budějovice, Czech Republic. [3]Present address: European Molecular Biology Laboratory, Heidelberg, Germany. [4]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands. [5]The Pirbright Institute, Surrey, UK. ✉email: petr.nguyen@prf.jcu.cz

understand major trends in karyotype evolution of moths and butterflies, are few[29,30] due to a lack of appropriate cytogenetic markers that can be used on this scale.

Mapping specific sequences on chromosome preparations by means of fluorescence in situ hybridization (FISH) allows us to identify particular chromosomes, study their potential rearrangements and origin, and their behaviour during cell divisions[29,31,32]. Various tandemly arrayed genes have been established as suitable markers for cytogenetic comparative studies. These universal markers have proved to be useful in a wide range of non-model species due to their conserved nature and ease of visualization by FISH methods[33–35]. The most commonly used markers are genes for major ribosomal RNAs (rDNA). Genes for 18S, 5.8S, and 28S ribosomal RNA form a transcription unit organized in clusters, which can contain hundreds or thousands of copies[36,37]. Genes for the 5S ribosomal RNA are also used[38–41]. 5S ribosomal RNA is distributed independently from the major rDNA array and is used as an independent marker. As with the major rDNA array, 5S rDNA can be localized in clusters containing tens to thousands of copies[38,42,43] but can also be found as singular copies scattered throughout the genome[43]. Abundant data on the number and localization of both 5S and the major rDNA gene clusters in animals and plants are available in public databases[44,45]. Finally, another group of markers used in cytogenetic studies includes the uridine-rich small nuclear RNA (U-rich snRNA) genes, which are an important part of the spliceosome. For cytogenetic purposes, U1 and U2 snRNA genes have been used. U1 snRNA gene clusters have been mapped in only a few species of Orthoptera[42,46], Isopoda[47], and fish[48]. U2 snRNA has been used only in a few fish species (e.g. Refs.[49–51]). U1 and U2 snRNA genes are relatively new markers often used in combination with other markers as major rDNA.

Despite their easy visualization and universality, rDNA and snRNA markers also have some limitations. Their evolution is highly dynamic, and changes in their distribution do not always reflect chromosome rearrangements[46,52]. They have been compared with mobile elements and in several cases have actually been found to be associated with transposons[48,53,54]. FISH experiments using 18S and 28S rDNA genes as probes successfully revealed concealed karyotype variation between populations and closely related species of both plants (e.g. Ref.[55]) and animals (e.g. Refs.[56,57]). Therefore, rDNA and snRNA genes might be good markers for chromosome evolution between closely related species or even intra-species evolution but are less informative with increasing evolutionary scale. To study such large-scale chromosome evolution patterns, additional markers should be developed that evolve less erratically.

Despite their great potential, histone genes have rarely been used in cytogenetic studies. Histone genes encode H1, H2A, H2B, H3, and H4 proteins, which have a strong affinity for DNA. Together, the histone proteins and DNA form a nucleosome, the basic unit of chromatin[58]. Histone genes usually form tandem arrays, as this facilitates efficient transcription[58,59]. The histone genes are conserved in their protein sequence and also in the distribution of their clusters in the genome[60,61]. This makes them ideal chromosomal markers[62] as differences in their number and position genuinely reflect chromosomal rearrangements[60]. Some examples of successful application of histone genes to fish[61], Bivalvia[63], and insects[34,39], including Lepidoptera[64,65], show their applicability in various organisms.

In this study, we analysed the chromosomal distribution of several universal cytogenetic markers, namely 18S and 5S rDNAs, U1 and U2 snRNA genes, and histone H3 genes, in 29 species of Lepidoptera to evaluate their applicability and resolution in the study of karyotype evolution. We found that some of the markers can be used successfully in all species, while others cannot be detected in certain species. To determine the reason for the unsuccessful detection of markers by FISH, we used quantitative PCR and Southern hybridization to estimate copy numbers and distribution patterns in different species. The obtained results provide not only information on the use of various markers in Lepidoptera, but also on trends in changes in the architecture of lepidopteran genomes.

Results

Localization of 18S rDNA and histone H3 genes. To visualize clusters of the major rRNA genes, we used FISH with a partial sequence of 18S rDNA from the codling moth, *Cydia pomonella* (Tortricidae) as a probe[25]. Since the nucleotide sequence of 18S rDNA is highly conserved, the probe successfully hybridized onto chromosomal preparations of all studied species sampled across the order Lepidoptera, as well as the representative of their sister order Trichoptera. Major rDNA clusters were detected at a terminal position in 22 out of 30 species. Only 6 species with interstitial clusters were documented. Multiple, up to 11, clusters were observed in approximately half of the studied species.

Although histone genes are known for their highly conserved protein sequence, they can differ significantly at the nucleotide level due to the degeneracy of codons. To ensure optimal hybridization, a fragment of the histone H3 gene was amplified, sequenced, and used as a specific probe from each species studied (Supplementary Table S1) except for few (for details see "Materials and methods"). To increase sensitivity of the FISH detection, we employed TSA-FISH which can detect unique sequences > 1300 bp[27]. In total, we successfully mapped the distribution of histone gene clusters in all studied species. In the vast majority, a single cluster was detected, located interstitially or terminally. Multiple histone clusters (2–3) were observed only in two lepidopteran species, *Tuta absoluta* and *Hyalophora cecropia*, and in the outgroup species *Glyphotaelius pellucidus* (Trichoptera).

All results from the mapping of 18S rDNA and histone H3 genes are summarized in Fig. 1 and Supplementary Table S2. For a complete overview we also added information on chromosome numbers and the distribution of 18S rDNA and histone H3 genes available to date in other Lepidoptera.

Trichoptera and non-Ditrysia. The diploid chromosome number of the caddis fly, *Glyphotaelius pellucidus* (Limnephiloidea) 2n♀ = 59, Z0/2n♂ = 60, ZZ, was described previously[71]. FISH experiments using 18S rDNA probe revealed a pair of terminal signals on one autosomal bivalent in this species (Supplementary

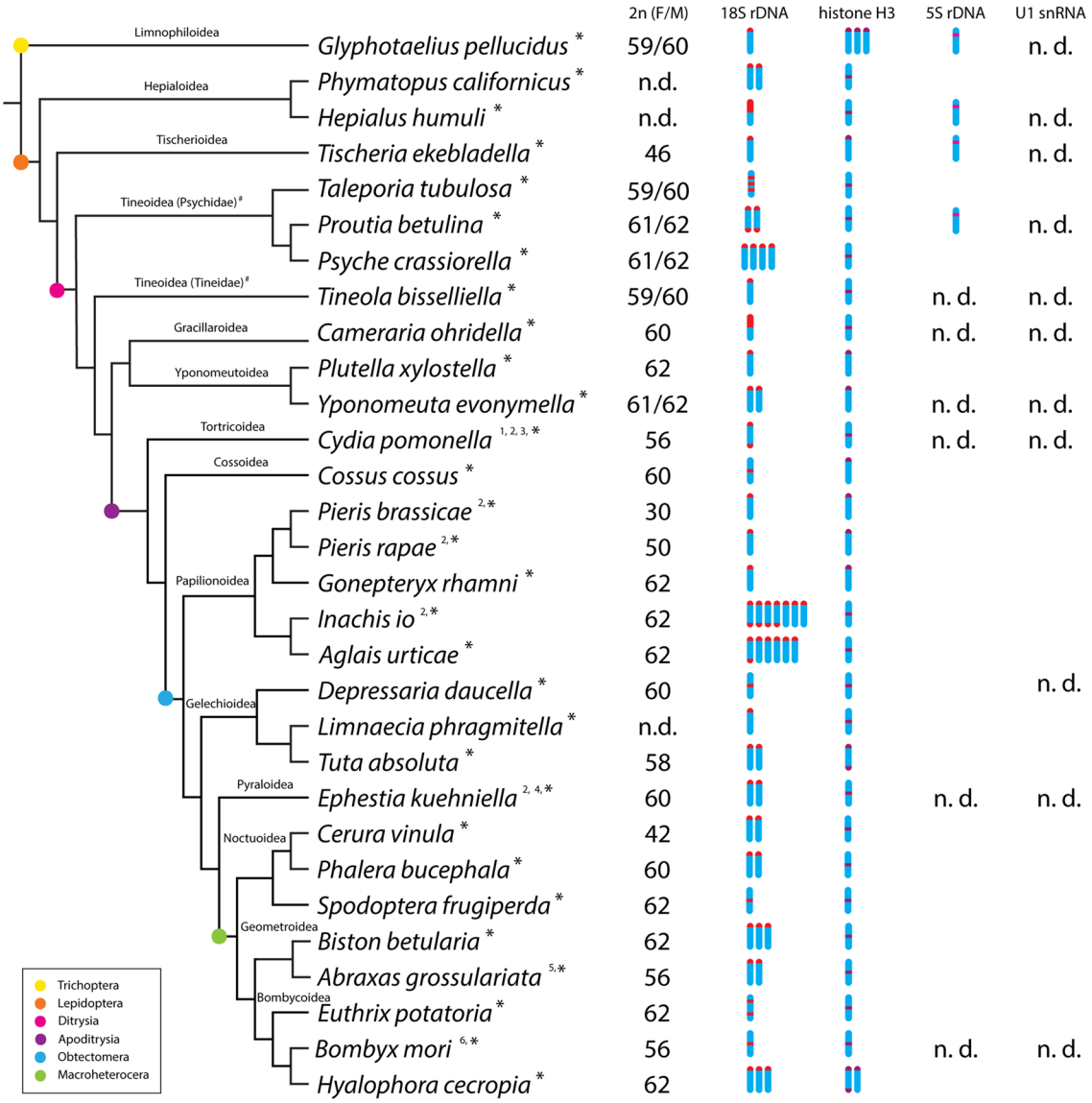

Figure 1. Overview of the number and position of 18S rDNA, histone H3, 5S rDNA, and U1 snRNA markers in haploid genomes of studied species. Phylogenetic relationships are based on Refs.[66–68]. #Tineoidea are considered paraphyletic. Data was obtained: *in this study; 1—[25], 18S rDNA; 2—[29], 18S rDNA; 3—[64] histone H3; 4—[69], 18S rDNA; 5—[70], 18S rDNA; 6—[31], 18S rDNA. n.d.—not detected. F/M—female and male diploid chromosome numbers, if different. A complete list of all species analysed so far for the distribution of studied markers, including their chromosomal numbers and references, is given in Supplementary Table S2. The figure was created in Adobe Illustrator 2020, version 24.0 (www.adobe.com).

Fig. S1a). Interestingly, various strong heterochromatin blocks in almost all chromosome bivalents were visible after staining with DAPI. Some of these heterochromatin patterns could potentially be used for chromosome identification. Hybridization of histone H3 probe revealed three terminal clusters of histone genes on three different bivalents (Supplementary Fig. S1b). This is one of three cases in our study where we observed multiple histone clusters (Fig. 1).

Two species of the superfamily Hepialoidea were examined, namely the ghost moth, *Hepialus humuli*, and the lupine ghost moth, *Phymatopus californicus*. Diploid chromosome numbers of these two species have not been described yet. Due to the lack of mitotic nuclei, we were not able to determine chromosomal numbers in this study. In *H. humuli*, the 18S rDNA probe highlighted approximately half of one pachytene bivalent (Supplementary Fig. S1c). Hybridization signals colocalized with a DAPI-positive heterochromatin block. In *P. californicus,* two chromosomal bivalents were detected, each bearing an rDNA cluster at the chromosome terminus (Supplementary Fig. S1e). The histone H3 probe revealed one bivalent with an interstitial cluster of histone genes in both *H. humuli* and *P. californicus* (Supplementary Fig. S1d, f).

The oak leaf miner, *Tischeria ekebladella,* was examined as a representative of the Tischeroidea superfamily. Its diploid chromosomal number 2n = 46 (for both sexes) was determined previously[72]. Indeed, n = 23 was confirmed in this study (Supplementary Fig. S2a). After DAPI staining, a strong heterochromatin block with terminal

or subterminal location was visible in three pachytene bivalents. The subterminal heterochromatin block was adjacent to a terminal rDNA cluster highlighted by the 18S rDNA probe on one of the longer bivalents in male pachytene nuclei (Supplementary Fig. S1g). A single histone gene cluster was localized at the end of another bivalent (Supplementary Fig. S1h).

Basal Ditrysia. Three bagworm species from the family Psychidae, namely *Taleporia tubulosa* (2n♀ = 59, Z0/2n♂ = 60, ZZ), *Proutia betulina* (2n♀ = 61, Z0/2n♂ = 62, ZZ), and *Psyche crassiorella* (2n♀ = 61, Z0/2n♂ = 62, ZZ)[73–75] and one species from the family Tineidae, the common clothes moth *Tineola bisselliella* (2n♀ = 59, Z0/2n♂ = 60, ZZ)[76], were studied. The 18S rDNA probe revealed a distinct hybridization pattern in each species. In pachytene nuclei of *T. tubulosa,* an extraordinary pattern of three strong interstitial rDNA signals located on a single bivalent with regular spacing was observed (Supplementary Fig. S3a). In pachytene nuclei of *P. betulina,* one bivalent with signals on both ends and two bivalents bearing one terminal signal each were observed (Supplementary Fig. S3c). In *P. crassiorella*, four terminal signals were located on four bivalents (Supplementary Fig. S3e). In *T. bisselliella* male pachytene nuclei, a single rDNA locus was detected in a subterminal region of a pachytene bivalent (Supplementary Fig. S3g). Only one interstitial cluster of histone genes was observed in all four species, *T. tubulosa*, *P. betulina*, *P. crassiorella*, and *T. bisselliella* (Supplementary Fig. S3b,d,f,h).

The diploid chromosome number 2n = 60 of the horse-chestnut leaf miner, *Cameraria ohridella* (Gracillarioidea), was determined previously[77]. The 18S rDNA probe hybridized to a terminal region of one bivalent in male pachytene nuclei (Supplementary Fig. S4a). A strong, yet discontinuous signal covered approximately one fourth of the bivalent in a pattern similar to the one observed in *H. humuli* (see above). Mapping of the histone H3 gene showed one interstitial histone cluster on the rDNA bearing bivalent (Supplementary Fig. S4b).

Two species from the superfamily Yponomeutoidea were examined, i.e. the diamondback moth *Plutella xylostella* (Plutellidae; 2n = 62[78]) and the bird-cherry ermine moth *Yponomeuta evonymella* (Yponomeutidae; 2n♀ = 61, Z_1Z_2W/2n♂ = 62, $Z_1Z_1Z_2Z_2$; Ref.[79] and references therein). FISH experiments carried out on pachytene nuclei of *P. xylostella* revealed a single terminal cluster of rDNA genes (Supplementary Fig. S4c). The histone H3 probe revealed one terminal cluster which colocalized with a strong heterochromatic block (Supplementary Fig. S4d). On chromosomal preparations of *Y. evonymella*, the 18S rDNA probe showed two bivalents with terminal signals of similar size (Supplementary Fig. S4e). One terminal histone cluster was observed in pachytene nuclei of *Y. evonymella* (Supplementary Fig. S4f).

Apoditrysia. In a representative of the Cossoidea superfamily, the goat moth *Cossus cossus*, we determined the diploid male chromosome number 2n = 60 (Supplementary Fig. S5a). FISH with the 18S rDNA probe on male pachytene nuclei revealed one chromosomal pair bearing an interstitial cluster which colocalized with a small block of DAPI-positive heterochromatin (Supplementary Fig. S5a). The histone H3 probe labelled one cluster at the end of one chromosome bivalent (Supplementary Fig. S5b). However, it should be noted that due to the lack of material, our FISH experiments were performed on only one male *C. cossus* larva. The karyotype of the codling moth, *Cydia pomonella* (Totricidae), was already described as 2n = 56 by Ref.[80] and later verified by Ref.[25]. Our results of mapping of the 18S rDNA and histone H3 probes (Supplementary Fig. S5c,d) confirmed previously published data, i.e. two rDNA clusters at both ends of a single chromosome bivalent and another bivalent bearing one interstitial histone cluster[25,64].

Obtectomera. Within the superfamily Papilionoidea we studied three species belonging to the family Pieridae and two species of the Nymphalidae family. The three studied pierids, namely the small cabbage white *Pieris rapae* (2n = 50[81]), the cabbage white *Pieris brassicae* (2n = 30[81]), and the common brimstone *Gonepteryx rhamni* (2n = 62[81]), differ in chromosomal numbers, however, mapping of the 18S rDNA and histone H3 genes revealed common hybridization patterns for both markers (Supplementary Fig. S6). Consistent with previous reports[29], we identified one bivalent bearing a terminal rDNA cluster in *P. rapae* (Supplementary Fig. S6a) and *P. brassicae* (Supplementary Fig. S6c). We also observed this pattern in autosome pair of *G. rhamni* (Supplementary Fig. S6e). In pachytene nuclei of *P. brassicae*, small DAPI-positive blocks of heterochromatin were observed at the ends of several bivalents (Supplementary Fig. S6c). In *G. rhamni,* only one block of heterochromatin was visible, which colocalized with the 18S rDNA signal (Supplementary Fig. S6e). The histone H3 probe highlighted the terminal region in one chromosome pair in all three species (Supplementary Fig. S6b,d,f). Moreover, histone bearing chromosomes clearly correspond to autosomes in *P brassicae*, in which the sex chromosome bivalent was identified by a typical pairing of W and Z chromosomes (Supplementary Fig. S6d).

From the family Nymphalidae, the small tortoiseshell *Aglais urticae* and the peacock butterfly *Inachis io* were examined. Both species have a chromosome number 2n = 62, reported previously[81] and confirmed in this study (Supplementary Fig. S2b,c). In male pachytene nuclei of *A. urticae*, six to seven small rDNA clusters were observed (Supplementary Fig. S7a). FISH with the histone H3 probe revealed one interstitial cluster colocalizing with a heterochromatin block (Supplementary Fig. S7b). Mapping of 18S rDNA genes in *I. io*, which was done previously[29], revealed up to 11 small terminal clusters in pachytene nuclei, three bivalents bearing one terminal signal and four bivalents carrying terminal signals at both ends. To increase the sensitivity of detection, we repeated this experiment using TSA-FISH. Our data confirm the previous identification and distribution of eleven 18S rDNA clusters in *I. io* (Supplementary Fig. S7c). Similar to *A. urticae*, we mapped a single histone cluster to an interstitial region of a bivalent, which colocalized with a block of heterochromatin (Supplementary Fig. S7d).

We studied three species from the superfamily Gelechioidea, namely the dingy flat body moth, *Depressaria daucella* (Depressariidae), the shy cosmet moth, *Limnaecia phragmitella* (Cosmopterigidae), and the tomato leafminer, *Tuta absoluta* (Gelechiidae). The diploid chromosome number of 2n = 60 in *D. daucella* was reported recently[22]. The 18S rDNA probe revealed one interstitial cluster of major rDNA (Supplementary Fig. S8a).

Similarly, we detected one interstitial histone cluster (Supplementary Fig. S8b). In *L. phragmitella*, the number of chromosomes was not determined previously, and we also failed to determine it due to the lack of mitotic chromosomes. However, using FISH mapping on pachytene chromosomes, we successfully identified one terminal rDNA cluster colocalized with a DAPI-positive block of heterochromatin (Supplementary Fig. S8c) and one interstitial cluster of histone genes (Supplementary Fig. S8d). In the *T. absoluta* strain used in this study, a diploid chromosome number of 2n = 58 was described previously[82] and confirmed in another study[22]. The 18S rDNA probe highlighted two clusters in terminal regions of two autosomal bivalents (Supplementary Fig. S8e). Histone gene clusters were detected at both ends of a pachytene bivalent (Supplementary Fig. S8f), which makes *T. absoluta* one of only two lepidopteran species with multiple histone gene clusters described so far.

The only representative of the superfamily Pyraloidea included in our study was the Mediterranean flour moth, *Ephestia kuehniella*. Its diploid chromosome number of 2n = 60 was described previously[83]. Two terminal rDNA clusters present on two chromosome bivalents were identified by Ref.[69], which was later confirmed by means of FISH[29]. To complete the dataset, we additionally mapped the histone H3 probe on male pachytene nuclei, which revealed one chromosome bivalent bearing a single interstitial cluster of histone genes colocalizing with a block of heterochromatin (Supplementary Fig. S10a).

Macroheterocera. Three members of the Notodontidae and Noctuidae families within the superfamily Noctuoidea were examined. In male pachytene complements of the puss moth, *Cerura vinula* (Notodontidae), we verified the diploid chromosome number of 2n = 42 (Supplementary Fig. S2d) previously reported[81] and observed two autosomal bivalents carrying a terminal rDNA cluster (Supplementary Fig. S9a). One interstitial cluster of histone genes was detected by the histone H3 probe (Supplementary Fig. S9b). In female pachytene nuclei of the buff-tip, *Phalera bucephala* (Notodontidae), with a diploid number of chromosomes 2n = 60[81], the same hybridization pattern for both markers as in *C. vinula* was observed (Supplementary Fig. S9c,d). In the fall armyworm, *Spodoptera frugiperda* (Noctuidae), with a diploid chromosome number of 2n = 62[81], only one bivalent bearing an interstitial cluster of rDNA genes was identified by FISH (Supplementary Fig. S9e). Similar to the other two species, one interstitial histone cluster was detected in one of the bivalents in male pachytene nuclei (Supplementary Fig. S9f).

Two representatives of the superfamily Geometroidea were included in our study, the peppered moth *Biston betularia* and the magpie moth *Abraxas grossulariata* (both Geometridae). In *A. grossulariata*, the diploid number of chromosomes 2n = 56 was reported in an earlier study[70], which also detected a single terminal rDNA cluster on W and Z sex chromosomes. On chromosomal preparations of *A. grossulariata,* we identified one interstitial cluster of histone genes (Supplementary Fig. S10b). Moreover, numerous strong DAPI-positive heterochromatin blocks were detected (Supplementary Fig. S10b), which is also in agreement with previous observations[70]. The diploid chromosome number of 2n = 62 was previously reported for *B. betularia*[84] and the same material was used in this study. In male pachytene nuclei, we identified three bivalents bearing a single small terminal rDNA cluster each (Supplementary Fig. S10c). Using the histone H3 probe, a single interstitial histone cluster was detected on one of the autosomal bivalents in female pachytene nuclei (Supplementary Fig. S10d).

Species from three different families were explored within the superfamily Bombycoidea, the drinker moth *Euthrix potatoria* (Lasiocampidae), the silkworm *Bombyx mori* (Bombycidae), and the cecropia silkmoth *Hyalophora cecropia* (Saturnidae). A diploid chromosome number of 2n = 62 was previously described in *E. potatoria*[81] and was confirmed by our results (Supplementary Fig. S2e). Hybridization of the 18S rDNA probe revealed an interesting distribution of rDNA genes, namely two interstitially located clusters within one pachytene bivalent (Supplementary Fig. S11a). The histone H3 probe uncovered one interstitial cluster of histone genes (Fig. S11b). The diploid karyotype of *B. mori* consists of 2n = 56 chromosomes[85]. Distribution of rDNA was previously reported as a single interstitial rDNA cluster[29]. We confirmed this in male pachytene preparations (Supplementary Fig. S11c). Moreover, the FISH experiments with the histone H3 probe also revealed an interstitial position of a single histone gene cluster in one of the chromosomal pairs (Supplementary Fig. S11d).

The karyotype of *H. cecropia* consists of 2n = 62 chromosomes, as previously reported[86], which is corroborated also by our observation (Supplementary Fig. S2f). Three terminal clusters of rDNA genes in three different bivalents were mapped by the 18S rDNA probe (Supplementary Fig. S11e). Histone H3 mapping revealed two histone gene clusters at both ends of one bivalent and another cluster at one end of another bivalent colocalizing with strong blocks of heterochromatin. Bivalents bearing the histone clusters were almost exclusively associated in pachytene complements forming a specific configuration (Supplementary Fig. S11f).

Mapping of 5S rDNA and U1 and U2 snRNA genes. The 5S rDNA gene and U1 and U2 snRNA genes have never been used as cytogenetic markers in the order Lepidoptera. Therefore we decided to test their suitability for comparative analysis within this order. We chose nine species from different families across the whole order Lepidoptera with a focus on basal groups, namely *H. humuli* (Hepialidae), *T. ekebladella* (Tischeriidae), *T. bisselliella* (Tineidae), *T. tubulosa* (Psychidae) *C. ohridella* (Gracillariidae), *Y. evonymella* (Yponomeutidae), *C. pomonella* (Tortricidae), *E. kuehniella* (Pyralidae), *B. mori* (Bombycidae), and one outgroup species, *G. pellucidus* (Limnephilidae), from the sister order Trichoptera. We amplified and labelled species-specific probes for the 5S rDNA, U1 and U2 snRNA genes, and used them in FISH experiments in the respective species (Supplementary Table S1).

Although we used an optimized TSA-FISH protocol to maximize sensitivity of the FISH experiments, we successfully mapped 5S rDNA only in four species, namely *G. pellucidus*, *H. humuli*, *T. ekebladella*, and *T. tubulosa*. In all four species, we detected one subterminal cluster of 5S rDNA genes on one chromosomal pair (Supplementary Fig. S12). In the other species, no clear hybridization signals were identified. The U1 and U2 snRNA genes did not show any hybridization signals in any of the ten species studied (summary of results in Fig. 1). The negative

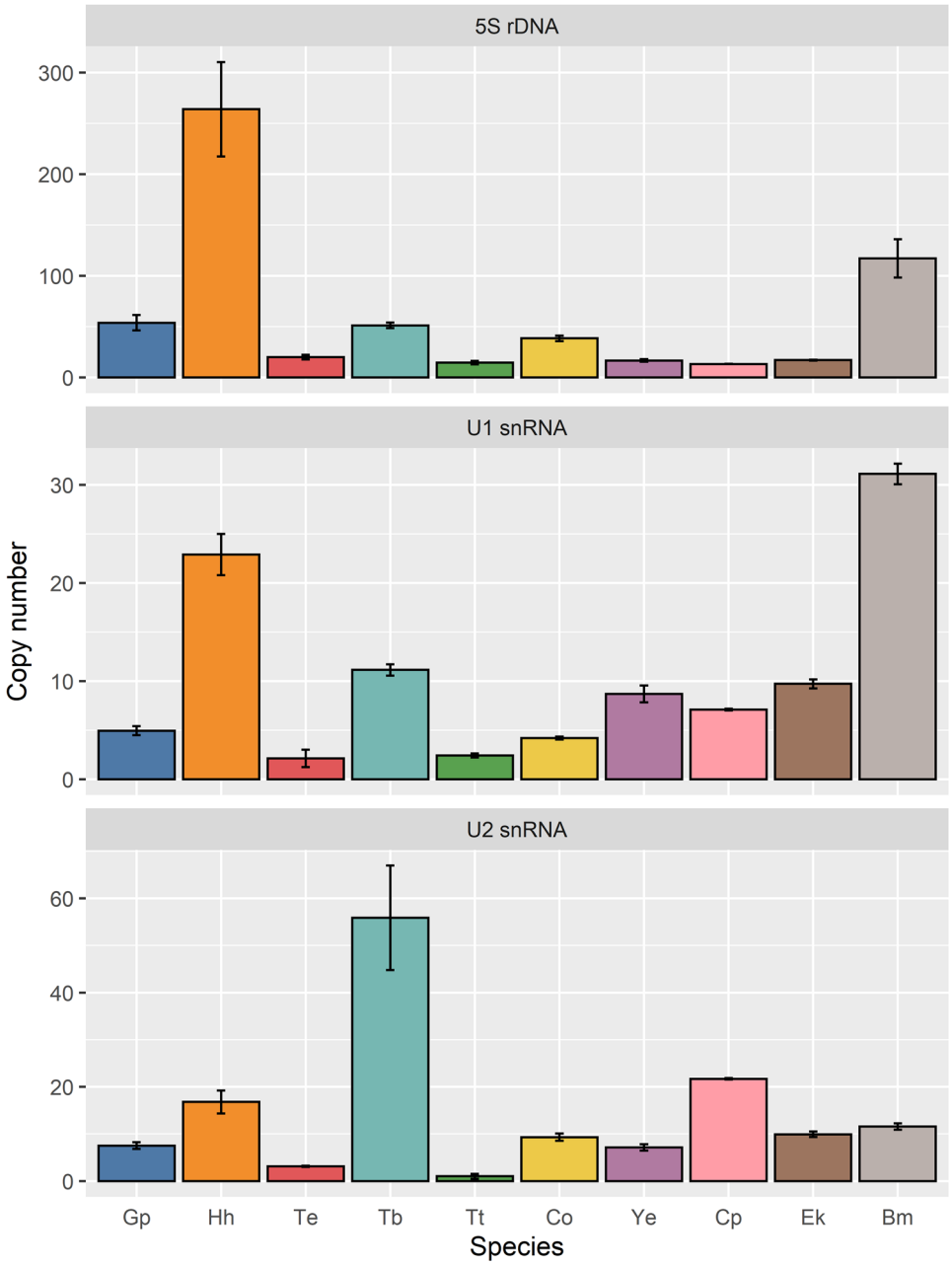

Figure 2. Estimated copy numbers of 5S rDNA, U1 snRNA, and U2 snRNA genes per haploid genome in selected species. Gp, *G. pellucidus* (Trichoptera, outgroup); Hh, *H. humuli*; Te, *T. ekebladella*; Tb, *T. bisselliella*; Tt, *T. tubulosa*; Co, *C. ohridella*; Ye, *Y. evonymella*; Cp, *C. pomonella*; Ek, *E. kuehniella*; Bm, *B. mori*. For a summary of the qPCR results, see Supplementary Table S3.

results of the FISH experiments suggest that the genomic arrangement of these three genes is not suitable for FISH mapping in Lepidoptera. For example, these genes may occur in low numbers in tandem arrays or may be scattered throughout the genome, rather than clustered. To test these hypotheses, we carried out quantitative PCR (qPCR) and Southern hybridization.

qPCR experiments. Quantitative PCR was carried out to estimate relative copy number of 5S rDNA, U1 snRNA, and U2 snRNA genes in ten representatives probed for these genes by TSA-FISH (see above). Estimated copy numbers are summarized in Fig. 2 and Supplementary Table S3. In the case of 5S rDNA, estimated copy numbers ranging from 13 to 264 copies (mean 60.602 SD ± 14.506, median 29.198). The 5S rDNA copy number was higher (> 50 copies) in three of the four species in which the 5S rDNA locus was detected by TSA-FISH, namely *G. pellucidus, H. humuli,* and *T. tubulosa*. Using TSA-FISH we also localized 5S rDNA in *T. ekebladella*, although its copy number was much lower (~ 20 copies) and comparable with other species in which 5S rDNA

could not be localized. However, we were unable to localize 5S rDNA by TSA-FISH in *C. ohridella* and *B. mori* with 5S rDNA copy numbers of 38 and > 100, respectively. For U1 snRNA and U2 snRNA, the results obtained showed low copy numbers in all species ranging from 2 to 31 copies (mean 10.436 SD ± 0.595, median 7.885) and from 1 to 56 copies (mean 14.387 SD ± 3.331, median 9.611) per haploid genome, respectively.

Southern hybridization. To test whether 5S rDNA and U1 snRNA genes are organized in tandem arrays, we performed Southern hybridization in ten selected species (see above). The U2 snRNA was excluded from this analysis due to difficulties in preparing digoxigenin-labelled probes.

Southern hybridization of the 5S rDNA probe was successful in all species tested (Supplementary Fig. S13). Results revealed multiple DNA fragments mostly > 2000 bp bearing the target sequence in all species examined. The intensity of hybridization signals was mostly uniform, although stronger bands correlating with multiple gene copies were identified e.g. in *G. pellucidus*, *H. humuli*, and *T. bisselliella*. Strong bands of smaller size, which presumably correspond to identical repeat units derived from tandem arrays, were observed only in *G. pellucidus*. Given the low copy numbers indicated by qPCR, the Southern hybridization results suggest the 5S rRNA gene copies are either scattered throughout the genome or loosely associated, i.e. individual copies are separated by varying spacers longer than 2000 bp.

Hybridization patterns using the U1 snRNA probe were similar to the 5S rDNA patterns in all species tested (Supplementary Fig. S14). Multiple bands of mostly weak intensity were observed, which implies that the 10 copies determined on average by qPCR are mostly either separated by long spacers or scattered across the genome in all species studied. Stronger bands corresponding to DNA fragments bearing multiple U1 snRNA copies were detected e.g. in *C. ohridella* and *C. pomonella* (Supplementary Fig. S14d,f). In the latter, however, the stronger bands can comprise multiple bands due to insufficient separation of long fragments (Supplementary Fig. S14f). In *T. ekebladella* and *T. tubulosa*, we were not able to successfully perform Southern hybridization, probably due to the low quality of the input gDNA and/or insufficiently labelled probes.

Discussion

In this study, we tested whether commonly used cytogenetic markers, namely 18S rDNA, histone H3, 5S rDNA, and U1 and U2 snRNA genes, are applicable and informative for studies of karyotype evolution in Lepidoptera. We employed fluorescence in situ hybridization techniques, rDNA-FISH and TSA-FISH, which enhance hybridization signals by antibody amplification and enzymatically catalysed reporter deposition, respectively. We complemented our FISH results by estimating the copy number of the markers by qPCR and characterizing their genomic organization using Southern hybridization.

Nguyen et al.[29] reviewed available data on the distribution of major rDNA in Lepidoptera and mapped 18S rDNA in 18 ditrysian species from 4 superfamilies (Pyraloidea, Bombycoidea, Papilionoidea, Noctuoidea). The results suggested that in karyotypes with one locus, rDNA was usually localized interstitially, whereas in karyotypes with two or more clusters, rDNA loci were detected at chromosome ends. It was hypothesized that rDNA can spread to terminal chromosome regions by ectopic recombination between subtelomeric repetitive sequences. However, missing data from non-ditrysian and early diverging ditrysian families did not allow inferring an ancestral rDNA distribution. To fill these gaps, we carried out FISH with the 18S rDNA probe in 27 moth and butterfly species with a special focus on early diverging taxa. We also investigated one trichopteran species as an outgroup.

The results of this study (summarized in the Fig. 1) suggest that one terminal cluster of rDNA genes is an ancestral state, as it is present in the outgroup and across all lepidopteran families. In species with multiple chromosomes bearing rDNA, these clusters are usually located terminally and there is a trend towards an increase in rDNA loci in Lepidoptera. Interestingly, the rDNA loci also multiplied in the early diverging ditrysian lineage Psychidae (Supplementary Fig. S3). The highest numbers of rDNA clusters, 11 and 7, were detected in two nymphalid species, *I. io* and *A. urticae*, respectively (Supplementary Fig. S7). Other karyotype features, such as chromosomal number n = 31 and presence of a single interstitial histone cluster (Supplementary Fig. S7), which are both considered ancestral traits (Refs.[72,87]; this study, see below), do not point to any large-scale chromosomal rearrangements in these nymphalids. Thus, the multiplication of rDNA clusters in nymphalids concurs with the ectopic recombination-driven spread of rDNA into new loci. Remarkably, multiple interstitial rDNA clusters present on a single bivalent were also documented in some species. Three interstitial rDNA clusters within a single bivalent were detected in *T. tubulosa* (Supplementary Fig. S3a) and two interstitial clusters within one bivalent were observed in *E. potatoria* (Bombycoidea) (Supplementary Fig. S11a). In both cases, the multiple clusters probably originated from intrachromosomal rearrangements such as inversions of a region containing part of the rDNA cluster (cf. Ref.[88,89]). In the case of *H. humuli* (Hepialoidea) (Supplementary Fig. S1c) and *C. ohridella* (Gracillarioidea) (Supplementary Fig. S4a), rDNA covers almost half of the chromosome. In addition, in *H. humuli* the rDNA cluster colocalizes with a strong heterochromatin block indicating the presence of repetitive sequences potentially associated with rDNA. More detailed research is needed to determine the mechanism of rDNA spread in these two species. Our data show that the multiplication of the major rDNA cluster occurs in multiple lepidopteran families and via different mechanisms, without any clear evolutionary pattern. This erratic behaviour makes the major rDNA an uninformative marker for the study of karyotype evolution in Lepidoptera.

Histone H3 genes have previously been mapped in several lepidopteran species[64,65,90]. One interstitial cluster of histone H3 genes was identified consistently in five species of the family Tortricidae[64]. Histone H3 genes were also localized in four *Leptidea* spp. (Pieridae)[65,90]. The position of the histone gene cluster was stable in *L. amurensis*, but in the other three *Leptidea* species, the number and position varied even among the offspring of one female. The karyotype evolution of *Leptidea* butterflies is known to be dynamic, characterized by unstable chromosome numbers[65], and the distribution of histone gene clusters thus reflects this instability[65,91]. To analyse

common trends in histone cluster repatterning across Lepidoptera, we mapped histone H3 genes in 29 moth and butterfly species and one caddisfly outgroup.

In the vast majority of species, TSA-FISH with the histone H3 probe revealed a single bivalent bearing the histone gene cluster (summarized in Fig. 1). This pattern was conserved in several superfamilies, such as Hepialoidea, Tineoidea, Geometroidea, Noctuoidea, and Bombycoidea. In the superfamily Papilionoidea, one interstitial histone gene cluster was observed in nymphalids (Supplementary Fig. S7), whereas in representatives of the family Pieridae, the cluster was identified at the terminal region of a bivalent (Supplementary Fig. S6). This difference in position is most likely the result of an inversion as no additional clusters were identified. This inversion can be one of many chromosomal rearrangements which seem to be typical for the genus *Pieris*[85,92]. A single terminal histone gene cluster was also characteristic of the superfamilies Tischeroidea, Yponomeutoidea, and Cossoidea, although more species need to be tested in these taxa. Multiple clusters were observed only in three species. The caddisfly *G. pellucidus* had three terminal clusters on different bivalents (Supplementary Fig. S1b), whereas *T. absoluta* (Gelechioidea) had two clusters on both ends of a single bivalent (Supplementary Fig. S8f). In *H. cecropia* (Bombycoidea), three terminal clusters were present on two bivalents (Supplementary Fig. S11f). Taken together, the ancestral state of histone genes is probably a single interstitially located cluster. In some taxa, the cluster moved to the chromosome end, allowing its further spread to terminal regions of the same or other chromosomes, probably due to ectopic recombination (cf. Ref.[29]). The localization of the histone gene cluster seems to be very conserved in Lepidoptera, with the exception of *Leptidea* spp.[65], and its changes indicate chromosomal rearrangements such as inversions, translocations, or chromosomal fusions and fissions (cf. Ref.[65]). Therefore, the histone H3 gene cluster is a good marker to study karyotype evolution in Lepidoptera.

Genes for 5S rRNA and U1 and U2 snRNAs have not yet been localized in lepidopteran genomes. Nine species sampled across Lepidoptera, namely *H. humuli* (Hepialidae), *T. ekebladella* (Tischeriidae), *T. bisselliella* (Tineidae), *T. tubulosa* (Psychidae) *C. ohridella* (Gracillariidae), *Y. evonymella* (Yponomeutidae), *C. pomonella* (Totricidae), *E. kuehniella* (Pyralidae), and *B. mori* (Bombycidae) were analysed along with a trichopteran outgroup, *G. pellucidus* (Limnephilidae). The genes and corresponding probes were very short (≤ 140 bp, Supplementary Table S4). Therefore, we used TSA-FISH, which allows the detection of single-copy genes ≥ 1300 bp[27]. Despite the optimization of the protocol, we were unable to localize the U1 and U2 snRNA genes in any of the species studied. The 5S rDNA clusters were detected only in the caddisfly *G. pellucidus* and representatives of early diverging lepidopteran lineages, namely *H. humuli*, *T. ekebladella*, and *T. tubulosa*. In all these species, TSA-FISH revealed a single interstitial 5S rDNA cluster (Fig. 1, Supplementary Fig. S12). It is tempting to speculate that the observed phylogenetic pattern could reflect a genome reorganization in Ditrysia, i.e. a lineage comprising 98% of extant moths and butterflies[17], in which *Hox* gene amplification occurred[93] and over 1000 novel gene families emerged[94]. However, more data on the distribution of 5S rDNA in early diverging lineages is needed to confirm whether this pattern is consistent.

To find out why 5S rDNA, U1 and U2 snRNAs were not detected by FISH, we determined the copy number of the genes and tested whether the gene copies are arranged in tandem. Quantitative PCR revealed that copy numbers of 5S rRNA genes vary greatly between species (Fig. 2). An upper limit of the 5S rRNA gene copy number was observed in *H. humuli*, which may correlate with its likely large genome size. Although the genome size of *H. humuli* is unknown, the C-value of other hepialids, *Thitarodes (Hepialus)* sp. and *Triodia sylvina*, is 2.92 Gb[95] and 1.8 Gb[96], respectively. However, copy number alone cannot explain the detectability of 5S rDNA in Lepidoptera. In *T. ekebladella*, we found approximately 20 copies of 5S rDNA genes, which we were able to detect by TSA-FISH, while we were not able to detect more than 100 copies of 5S rDNA genes in *B. mori*. In all species examined, the total length of the 5S rDNA cluster should be above the detection threshold of 1300 bp[27], if all the copies are arranged in tandem. However, results of Southern hybridization revealed multiple bands with fragment length > 2000 bp in all species (Supplementary Fig. S13), which suggests that the gene copies are scattered throughout the genome. Indeed, Vierna et al.[43] reported the presence of ten 5S rDNA clusters in the *B. mori* genome based on the analysis of genomic data. Alternatively, the copies can be only loosely clustered, i.e. separated by long spacers varying both in size and sequence. Clusters of 5S rDNA genes have been successfully mapped in many taxa[38,42,97,98]. However, the presence of multiple loci, which remain undetected even by TSA-FISH, questions homology of detected clusters and the usefulness of this marker in studies on karyotype evolution.

Copy number estimates for U1 and U2 snRNA genes by qPCR revealed much lower numbers than for 5S rDNA genes. These differences in copy number between 5S and U1 and U2 snRNAs seem to be consistent in Metazoa[87,99–101]. Based on our Southern hybridization results, the organization of U1 snRNA copies was quite similar to 5S rDNA, as multiple long fragments bearing the studied genes were observed (Supplementary Fig. S14). This means that successful detection of U1 and U2 snRNA clusters by FISH in some taxa[35,46,48,102] is the exception rather than the rule, and these genes are not universally applicable cytogenetic markers.

Taken together, 5S rDNA, U1 and U2 snRNA genes are not suitable markers for comparative cytogenetic studies in Lepidoptera. With a few exceptions, no clear cluster organization was detected by in situ hybridization. Their scattered organization and/or the presence of long spacer sequences between the genes does not allow for the observation of specific hybridization patterns and thereby the reconstruction of karyotype evolution. On the contrary, hybridization of 18S rDNA and histone H3 genes revealed a clustered organization of these genes in all species studied. Mapping of 18S rDNA showed rather dynamic evolution of the major rDNA, which does not always reflect chromosomal changes. However, various patterns, numbers, and locations of rDNA clusters could provide information on the evolution of repetitive sequences in lepidopteran genomes. Even though the mapping of histone H3 genes requires a species-specific probe preparation, hybridization patterns seem to genuinely reflect chromosomal rearrangements that occurred during the evolution of lepidopteran species. Our study shows that the evaluation of cytogenetic markers can significantly contribute to research focused on comparative cytogenetics and evolutionary genetics not only in Lepidoptera, but in all eukaryotic species.

Material and methods

Insects. Examined lepidopteran species and one representative of caddisflies (Trichoptera), which was used as an outgroup, were either collected in the field or obtained from laboratory stocks. Some species were dissected immediately after collection. In the other species, captured females were left to lay eggs in plastic containers with host plants. Hatched larvae were then reared on their host plants or artificial diet. For a list of studied species, their origin, and details of rearing see Table S5.

Chromosome preparations. Meiotic and mitotic chromosomes from all the studied species were obtained from female and male gonads of 4th or 5th instar larvae. The only exception was *Gonepteryx rhamni*, which was dissected as young imago. Chromosomal preparations were made by spreading technique as described previously[103]. Briefly, dissections were performed in physiological solution[104]. The dissected gonads were hypotonized for 10 min (0.075 M KCl) and fixed in Carnoy's fixative (ethanol, chloroform, acetic acid; 6:3:1) for 15 min. They were then dissociated using tungsten needles in a drop of 60% acetic acid on a slide and spread using a heating plate set at 45 °C. Chromosome preparations were passed through an ethanol series (70%, 80% and 100% ethanol; 30–60 s each) and stored at − 20 °C or − 80 °C until further use.

FISH with 18S rDNA probe. A partial sequence of 18S rDNA was generated by PCR from male genomic DNA (gDNA) of the codling moth, *Cydia pomonella*, using a pair of specific primers as described previously[25] (Supplementary Table S4). This fragment was ligated into Promega pGem T-Easy Vector (Promega, Madison, WI, USA), cloned, purified by NucleoSpinPlasmid kit (Macherey-Nagel, Düren, Germany), verified by sequencing (SEQme, Dobříš, Czech Republic), and reamplified by PCR from plasmid. The reamplified 18S rDNA fragment was purified by the Wizard SV Gel and PCR Clean-Up System (Promega) and labelled by nick translation using Nick Translation Kit (Abbott Molecular Inc., Des Plaines, IL, USA) for 105 min at 15 °C. The 25 µL labelling reaction contained 500 ng DNA, 40 µM dATP, 40 µM dCTP, 40 µM dGTP, 14.4 µM dTTP, and 25.6 µM biotin-16-dUTP (Roche Diagnostics GmbH, Mannheim, Germany).

FISH experiments were carried out according to the previous study[25] with some modifications. Briefly, chromosome preparations were removed from the freezer, dehydrated in ethanol series, and air-dried. Preparations were treated with 100 µg/mL RNase A for 1 h at 37 °C to remove RNA and subsequently blocked in 5 × Denhardt's solution for 30 min at 37 °C. In the next step, the slides were denatured in 70% formamide in 2 × SSC for 3.5 min at 68 °C. After denaturation for 5 min at 90 °C, a probe mixture containing 25 ng of biotin-labelled 18S rDNA probe, 25 µg of sonicated salmon sperm, 50% deionized formamide, 10% dextran sulphate in 2 × SSC in a total volume of 10 µL was applied to the slide and hybridized overnight at 37 °C. The biotin-labelled probe was detected by Cy3-conjugated streptavidin (diluted 1:1000 with blocking solution) (Jackson ImmunoRes. Labs. Inc, West Grove, PA, USA). Signals were amplified with biotinylated anti-streptavidin (diluted 1:25 with blocking solution) (Vector Labs. Inc, Burlingame, CA, USA), which was again detected by Cy3-conjugated streptavidin (diluted 1:1000 with blocking solution). The preparations were counterstained with 0.5 µg/mL of DAPI (4′,6-diamidino-2-phenylindole) and mounted in antifade containing DABCO (1,4-diazabicyclo[2.2.2]octane).

FISH with tyramide signal amplification (TSA-FISH). To obtain specific histone H3, 5S rDNA, and U1 and U2 snRNA probes for each species or family, fragments of the respective genes were amplified by PCR using degenerate primers (Supplementary Table S4) and gDNA of each individual species as a template, as detailed previously[64]. Species-specific amplified gene fragments were cloned and verified by sequencing (SEQme) (Supplementary Table S1). The verified plasmids were purified by NucleoSpin Plasmid kit (Macherey–Nagel) and used as template DNA to prepare a labelled probe by PCR. Each 25 µL labelling reaction contained 1–10 ng template DNA, 1 × Ex *Taq* buffer, 1 mM each dATP, dCTP, and dGTP; 0.36 mM dTTP; 0.64 mM of fluorescein-12-dUTP (PerkinElmer, Waltham, MA, USA), 5 µmol of each primer, and 0.25 U TaKaRa Ex *Taq* DNA polymerase (TaKaRa, Otsu, Japan). Labelled probes were purified using Sephadex (Illustra Sephadex G-50 fine DNA grade). In *Inachis io* and *Tuta absoluta,* rDNA clusters were also mapped by TSA FISH with 18S rDNA probe labelled by fluorescein instead of FISH with biotin-labelled probe described above. Species-specific probes were generated for most species, except for *Phymatopus californicus*, to which we hybridized a probe from *Hepialus humuli*, *Psyche crassiorella* with a probe from *Taleporia tubulosa*, and *Pieris brassicae* with a probe from *Pieris rapae*.

TSA-FISH was performed according to the published protocol[27] with some modifications. Briefly, frozen chromosome preparations were dehydrated using an ethanol series. After drying, slides were treated with 10 mM HCl for 10 min at 37 °C to remove cytoplasm and incubated in 1% hydrogen peroxide for 30 min at room temperature to quench endogenous peroxidase activity. Preparations were digested with 100 µg/mL RNase A for 1 h at 37 °C and blocked with 5 × Denhardt's solution for 30 min at 37 °C. Thereafter, a 50 µL probe mixture containing 10–30 ng of labelled specific probe in 50% deionized formamide and 10% dextran sulfate in 2 × SSC was added to the slide, and the probe and chromosomes were simultaneously denatured for 5 min at 70 °C. Hybridization took place overnight at 37 °C. Hybridization signals were enhanced by Antifluorescein-HRP conjugate (PerkinElmer) diluted 1:1000 and incubated with tyramide solution (TSA Plus Fluorescein system, PerkinElmer) for 10–15 min for 5S rDNA and U1 and U2 snRNA and 5–7 min for histone H3. The preparations were counterstained and mounted in antifade containing DABCO with 0.5 µg/mL of DAPI.

Microscopy and image processing. Observation of chromosome preparations from FISH experiments was performed with a Zeiss Axioplan 2 microscope (Carl Zeiss, Jena, Germany) equipped with appropriate fluorescence filter sets. An Olympus CCD monochrome camera XM10 equipped with cellSens 1.9 digital imaging software (Olympus Europa Holding, Hamburg, Germany) was used to record and capture black-and-white

pictures. Images were captured separately for each fluorescent dye and then pseudocoloured and superimposed with Adobe Photoshop CS4, version 11.0.

Quantitative analysis of gene doses. Quantitative PCR (qPCR) was used to estimate relative copy numbers of three target genes, namely U1 and U2 snRNA and 5S rDNA, in *Glyphotaelius pellucidus*, *Hepialus humuli*, *Tischeria ekebladella*, *Taleporia tubulosa, Tineola bisselliella, Cameraria ohridella*, *Yponomeuta evonymella, Ephestia kuehniella*, *Cydia pomonella*, and *Bombyx mori*[19,76]. By comparing the genes of interest to a single-copy autosomal reference gene (*Acetylcholinesterase 2*, *Ace2*), their relative copy numbers were estimated based on a target to reference gene dose ratio formula (Ref.[105]; see below). The reference gene and genes of interest were analysed simultaneously in technical triplicates of three independent biological replicas. Due to small body size of some species, namely *T. ekebladella*, *T. tubulosa, T. bisselliella*, and *C. ohridella*, 5–10 individuals were pooled for gDNA extraction carried out using NucleoSpin Tissue kit (Macherey–Nagel), DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), or NucleoSpin DNA Insect kit (Macherey–Nagel) following manufacturer's instructions. One individual per biological replica was used for the other species.

The qPCR contained 1–10 ng of gDNA, an optimized concentration of primers per species (details in Supplementary Table S6) and Xceed qPCR SG Mix Lo-ROX (Institute of Applied Biotechnologies, Prague, Czech Republic) in a total volume of 10 µL. Amplification efficiencies (*E*) for each gene and species were determined by 0×, 5×, 25×, and 125× dilutions of pooled gDNA of all biological replicas. For all three markers in *C. pomonella* and 5S rDNA in *B. mori* SYBR Premix Ex *Taq* II Perfect Real Time (1×; TaKaRa) was used and amplification efficiencies were determined by 0×, 10×, 100× and 1000× dilutions (details in Supplementary Table S6). The experiments were carried out using the C1000 Thermal cycler CFX96 Real-Time System (Bio-Rad, Hercules, CA, USA) and data were analysed using software Bio-Rad CFX Manager 3.1. The target to reference gene dose ratio was calculated for each biological sample according the formula $R = [(1 + E_{\text{Reference}})^{\text{CtReference}}]/[(1 + E_{\text{Target}})^{\text{CtTarget}}]$, where *R* is a relative copy number of target gene, *E* is the primer efficiency and Ct = cycle threshold[105].

Southern hybridization. Southern hybridization was performed to independently estimate the copy number of U1 snRNA and 5S rDNA and to test whether the genes are tandemly arranged in the genomes of ten selected species. Cloned fragments of studied genes were reamplified by PCR using degenerate primers (Supplementary Table S4) and the products were used as template for labelling with digoxigenin-11-dUTP (Roche Diagnostics GmbH). Labelling and purification of the probes were done as for TSA-FISH probes (see above).

High-molecular-weight gDNA of the studied species was extracted by standard phenol–chloroform[106] or by cetyltrimethylammonium bromide (CTAB)[107] extraction. Three pairs of restriction enzymes with no restriction sites within the target sequences were selected (Supplementary Table S7) and digestion of gDNA was carried out overnight at 37 °C. Enzymes were inactivated by addition of loading buffer (50% glycerol, 250 mM EDTA, 5.9 mM bromophenol blue) or Gel Loading Dye, Purple (6×) (New England Biolabs, Ipswich, MA, USA) (for details see Supplementary Table S7). Five micrograms of digested DNA per well was separated using a 1% agarose gel in 1 × TBE buffer by horizontal electrophoresis at 5 V/cm. Southern hybridizations were carried out according to the published protocol[108] with some modifications. Briefly, after electrophoretic separation, DNA was denatured and transferred onto an Amersham Hybond-N + nylon membrane (GE Healthcare, Buckinghamshire, UK) by capillary flow. Hybridization of labelled probes (100 ng) was done overnight at 42 °C and the stringent washes on the subsequent day were performed at 68 °C. Probes were detected using Anti-Digoxigenin-AP (75 mU/mL; Roche Diagnostics GmbH) incubated with CDP-*Star* ready-to-use (Roche Diagnostics GmbH). Resulting chemiluminescence was recorded with a LAS-3000 Lumi-Imager (Fuji Photo Film Europe GmbH, Düsseldorf, Germany).

Data availability

All data generated or analysed during this study are included in this published article [and its Supplementary Information files]. Partial sequences of genes under study were deposited in GenBank under the acc. nos. MW149037–MW149046, MW194851–MW194870, and MW558903–MW558929.

Received: 26 March 2021; Accepted: 24 May 2021
Published online: 09 June 2021

References

1. Traut, W., Ahola, V., Smith, D. A. S., Gordon, I. J. & ffrench-Constant, R. H. Karyotypes versus genomes: The nymphalid butterflies *Melitaea cinxia*, *Danaus plexippus*, and *D. chrysippus*. *Cytogenet. Genome Res.* **153**, 46–53 (2017).
2. Heng, H. H. *et al.* A postgenomic perspective on molecular cytogenetics. *CG* **19**, 227–239 (2018).
3. Deakin, J. E. *et al.* Chromosomics: Bridging the gap between genomes and chromosomes. *Genes* **10**, 627. https://doi.org/10.3390/genes10080627 (2019).
4. Lewin, H. A. *et al.* Earth BioGenome Project: Sequencing life for the future of life. *PNAS* **115**, 4325–4333 (2018).
5. Claussen, U. Chromosomics. *Cytogenet. Genome Res.* **111**, 101–106 (2005).
6. Graphodatsky, A. S. Comparative chromosomics. *Mol. Biol.* **41**, 361–375 (2007).
7. Dutrillaux, B., Couturier, J., Richer, C.-L. & Viegas-Péquignot, E. Sequence of DNA replication in 277 R- and Q-bands of human chromosomes using a BrdU treatment. *Chromosoma* **58**, 51–61 (1976).
8. Bickmore, W. A. Karyotype analysis and chromosome banding. In *Encyclopedia of Life Sciences* (ed. Wiley) a0001160 (Wiley, 2001).
9. King, M. C-Banding studies on Australian hylid frogs: Secondary constriction structure and the concept of euchromatin transformation. *Chromosoma* **80**, 191–217 (1980).
10. Steiniger, G. E. & Mukherjee, A. B. Insect chromosome banding: technique for G- and Q-banding pattern in the mosquito *Aedes albopictus*. *Can. J. Genet. Cytol.* **17**, 241–244 (1975).
11. Brum-Zorrilla, N. & Postiglioni, A. Karyological studies on Uruguayan spiders I. Banding pattern in chromosomes of *Lycosa* species (Araneae-Lycosidae). *Genetica* **54**, 149–153 (1980).

12. Cabrero, J. & Camacho, J. P. M. Cytogenetic studies in gomphocerine grasshoppers. I. Comparative analysis of chromosome C-banding pattern. *Heredity* **56**, 365–372 (1986).
13. Chen, R., Song, W., Li, X. & An, Z. Chromosome G-banding in plants by inducing with trypsin and urea. *Cell Res.* **4**, 79–87 (1994).
14. Song, Y. C. *et al.* Comparisons of G-banding patterns in six species of the Poaceae. *Hereditas* **121**, 31–38 (2004).
15. Bedo, D. G. Karyotypic and chromosome banding studies of the potato tuber moth, *Phthorimaea operculella* (Zeller) (Lepidoptera, Gelechiidae). *Can. J. Genet. Cytol.* **26**, 141–145 (1984).
16. Prins, J. D. & Saitoh, K. Karyology and sex determination. In *Band 4: Arthropoda, 2 Hälfte: Insecta, Lepidoptera, Moths and Butterflies, Teilband/Part 36, Vol 2: Morphology, Physiology, and Development* (ed. Kükenthal, W.) (DE GRUYTER, 2003).
17. Van Nieukerken, E. J. *et al.* Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa* **3148**, 212. https://doi.org/10.11646/zootaxa.3148.1.41 (2011).
18. Yoshido, A. *et al.* Evolution of multiple sex-chromosomes associated with dynamic genome reshuffling in Leptidea wood-white butterflies. *Heredity* **125**, 138–154 (2020).
19. Nguyen, P. *et al.* Neo-sex chromosomes and adaptive potential in tortricid pests. *PNAS* **110**, 6931–6936 (2013).
20. Picq, S. *et al.* Insights into the structure of the spruce budworm (*Choristoneura fumiferana*) genome, as revealed by molecular cytogenetic analyses and a high-density linkage map. *G3* **8**, 2539–2549 (2018).
21. Asser-Kaiser, S. *et al.* Rapid emergence of baculovirus resistance in codling moth due to dominant, sex-linked inheritance. *Science* **317**, 1916–1918 (2007).
22. Carabajal Paladino, L. Z. *et al.* Sex chromosome turnover in moths of the diverse superfamily Gelechioidea. *Genome Biol. Evol.* **11**, 1307–1319 (2019).
23. Labbé, R., Caveney, S. & Donly, C. Genetic analysis of the xenobiotic resistance-associated ABC gene subfamilies of the Lepidoptera. *Insect Mol. Biol.* **20**, 243–256 (2010).
24. Wolf-I, K. W., Novák, K. & Marec, F. Kinetic organization of metaphase I bivalents in spermatogenesis of Lepidoptera and Trichoptera species with small chromosome numbers. *Heredity* **79**, 35–143 (1997).
25. Fuková, I., Nguyen, P. & Marec, F. Codling moth cytogenetics: Karyotype, chromosomal location of rDNA, and molecular differentiation of sex chromosomes. *Genome* **48**, 1083–1092 (2005).
26. Gan, Y. *et al.* Individual chromosome identification, chromosomal collinearity and genetic-physical integrated map in *Gossypium darwinii* and four D genome cotton species revealed by BAC-FISH. *Genes Genet. Syst.* **87**, 233–241 (2012).
27. Carabajal Paladino, L. Z., Nguyen, P., Šíchová, J. & Marec, F. Mapping of single-copy genes by TSA-FISH in the codling moth, *Cydia pomonella*. *BMC Genet.* **15**, S15. https://doi.org/10.1186/1471-2156-15-S2-S15 (2014).
28. Traut, W. Pachytene mapping in the female silkworm, *Bombyx mori* L. (Lepidoptera). *Chromosoma* **58**, 275–284 (1976).
29. Nguyen, P., Sahara, K., Yoshido, A. & Marec, F. Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). *Genetica* **138**, 343–354 (2010).
30. Vershinina, A. O., Anokhin, B. A. & Lukhtanov, V. A. Ribosomal DNA clusters and telomeric (TTAGG)n repeats in blue butterflies (Lepidoptera, Lycaenidae) with low and high chromosome numbers. *CCG* **9**, 161–171 (2015).
31. Yoshido, A., Bando, H., Yasukochi, Y. & Sahara, K. The *Bombyx mori* karyotype and the assignment of linkage groups. *Genetics* **170**, 675–685 (2005).
32. Lukhtanov, V. A. *et al.* Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. *PNAS* **115**, E9610–E9619 (2018).
33. Lee, Y.-I. *et al.* Satellite DNA in *Paphiopedilum* subgenus *Parvisepalum* as revealed by high-throughput sequencing and fluorescent in situ hybridization. *BMC Genom.* **19**, 578. https://doi.org/10.1186/s12864-018-4956-7 (2018).
34. Cabral-de-Mello, D. C., Martins, C., Souza, M. J. & Moura, R. C. Cytogenetic mapping of 5S and 18S rRNAs and H3 histone genes in 4 ancient Proscopiidae grasshopper species: Contribution to understanding the evolutionary dynamics of multigene families. *Cytogenet. Genome Res.* **132**, 89–93 (2011).
35. García-Souto, D., Troncoso, T., Pérez, M. & Pasantes, J. J. Molecular cytogenetic analysis of the European hake *Merluccius merluccius* (Merlucciidae, Gadiformes): U1 and U2 snRNA gene clusters map to the same location. *PLoS One* **10**, e0146150. https://doi.org/10.1371/journal.pone.0146150 (2015).
36. Kobayashi, T., Heck, D. J., Nomura, M. & Horiuchi, T. Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes Dev.* **12**, 3821–3830 (1998).
37. Prokopowich, C. D., Gregory, T. R. & Crease, T. J. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46**, 48–50 (2003).
38. Cabral-de-Mello, D. C., Moura, R. C. & Martins, C. Chromosomal mapping of repetitive DNAs in the beetle *Dichotomius geminatus* provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA similarity between the B chromosome and A complement. *Heredity* **104**, 393–400 (2010).
39. Cabral-de-Mello, D. C., Oliveira, S. G., de Moura, R. C. & Martins, C. Chromosomal organization of the 18S and 5S rRNAs and histone H3 genes in *Scarabaeinae coleopterans*: Insights into the evolutionary dynamics of multigene families and heterochromatin. *BMC Genet.* **12**, 88. https://doi.org/10.1186/1471-2156-12-88 (2011).
40. Badaeva, E. D. *et al.* A set of cytogenetic markers allows the precise identification of all A-genome chromosomes in diploid and polyploid wheat. *Cytogenet. Genome Res.* **146**, 71–79 (2015).
41. Wang, W. *et al.* Remarkable variation of ribosomal DNA organization and copy number in gnetophytes, a distinct lineage of gymnosperms. *Ann. Bot.* **123**, 767–781 (2019).
42. Palacios-Gimenez, O. M., Castillo, E. R., Martí, D. A. & Cabral-de-Mello, D. C. Tracking the evolution of sex chromosome systems in Melanoplinae grasshoppers through chromosomal mapping of repetitive DNA sequences. *BMC Evol. Biol.* **13**, 167. https://doi.org/10.1186/1471-2148-13-167 (2013).
43. Vierna, J., Wehner, S., Höner zu Siederdissen, C., Martínez-Lage, A. & Marz, M. Systematic analysis and evolution of 5S ribosomal DNA in metazoans. *Heredity* **111**, 410–421 (2013).
44. Garcia, S., Garnatje, T. & Kovařík, A. Plant rDNA database: Ribosomal DNA loci information goes online. *Chromosoma* **121**, 389–394 (2012).
45. Sochorová, J., Garcia, S., Gálvez, F., Symonová, R. & Kovařík, A. Evolutionary trends in animal ribosomal DNA loci: Introduction to a new online database. *Chromosoma* **127**, 141–150 (2018).
46. Anjos, A. *et al.* U1 snDNA clusters in grasshoppers: Chromosomal dynamics and genomic organization. *Heredity* **114**, 207–219 (2015).
47. Barzotti, R., Pelliccia, F. & Rocchi, A. Identifcation and characterization of U1 small nuclear RNA genes from two crustacean isopod species. *Chromosome Res.* **11**, 365–373 (2003).
48. Cabral-de-Mello, D. C., Valente, G. T., Nakajima, R. T. & Martins, C. Genomic organization and comparative chromosome mapping of the U1 snRNA gene in cichlid fish, with an emphasis in *Oreochromis niloticus*. *Chromosome Res.* **20**, 279–292 (2012).
49. Merlo, M. A. *et al.* Analysis of the histone cluster in Senegalese sole (*Solea senegalensis*): Evidence for a divergent evolution of two canonical histone clusters. *Genome* **60**, 441–453 (2017).

50. Piscor, D., Fernandes, C. A. & Parise-Maltempi, P. P. Conserved number of U2 snDNA sites in *Piabina argentea*, *Piabarchus stramineus* and two *Bryconamericus* species (Characidae, Stevardiinae). *Neotrop. Ichthyol.* **16**, e170066. https://doi.org/10.1590/1982-0224-20170066 (2018).
51. Sember, A. *et al.* Dynamics of tandemly repeated DNA sequences during evolution of diploid and tetraploid botiid loaches (Teleostei: Cobitoidea: Botiidae). *PLoS One* **13**, e0195054. https://doi.org/10.1371/journal.pone.0195054 (2018).
52. Poletto, A. B. *et al.* Chromosome differentiation patterns during cichlid fish evolution. *BMC Genet.* **11**, 50. https://doi.org/10.1186/1471-2156-11-50 (2010).
53. Symonová, R. *et al.* Genome differentiation in a species pair of coregonine fishes: An extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol. Biol.* **13**, 42. https://doi.org/10.1186/1471-2148-13-42 (2013).
54. de Sene, V. F. *et al.* Mapping of the retrotransposable elements *Rex1* and *Rex3* in chromosomes of *Eigenmannia* (Teleostei, Gymnotiformes, Sternopygidae). *Cytogenet. Genome Res.* **146**, 319–324 (2015).
55. Cai, Q., Zhang, D., Liu, Z.-L. & Wang, X.-R. Chromosomal localization of 5S and 18S rDNA in five species of subgenus *Strobus* and their implications for genome evolution of *Pinus*. *Ann. Bot.* **97**, 715–722 (2006).
56. Grozeva, S., Kuznetsova, V. & Anokhin, B. Karyotypes, male meiosis and comparative FISH mapping of 18S ribosomal DNA and telomeric $(TTAGG)_n$ repeat in eight species of true bugs (Hemiptera, Heteroptera). *CCG* **5**, 355–374 (2011).
57. Chirino, M. G. & Bressa, M. J. Karyotype evolution in progress: A new diploid number in *Belostoma candidulum* (Heteroptera: Belostomatidae) from Argentina leading to new insights into its ecology and evolution. *Eur. J. Entomol.* **111**, 165–174 (2014).
58. Roehrdanz, R., Heilmann, L., Senechal, P., Sears, S. & Evenson, P. Histone and ribosomal RNA repetitive gene clusters of the boll weevil are linked in a tandem array: Histone-ribosomal DNA repeats. *Insect Mol. Biol.* **19**, 463–471 (2010).
59. Schienman, J. E., Lozovskaya, E. R. & Strausbaugh, L. D. *Drosophila virilis* has atypical kinds and arrangements of histone repeats. *Chromosoma* **107**, 529–539 (1998).
60. Zhang, L., Bao, Z., Wang, S., Huang, X. & Hu, J. Chromosome rearrangements in Pectinidae (Bivalvia: Pteriomorphia) implied based on chromosomal localization of histone H3 gene in four scallops. *Genetica* **130**, 193–198 (2007).
61. Silva, D. M. Z. A. *et al.* Chromosomal organization of repetitive DNA sequences in *Astyanax bockmanni* (Teleostei, Characiformes): Dispersive location, association and co-localization in the genome. *Genetica* **141**, 329–336 (2013).
62. Maxson, R., Cohn, R., Kedes, L. & Mohun, T. Expression and organization of histone genes. *Annu. Rev. Genet.* **17**, 239–277 (1983).
63. Eirín-López, J. M. *et al.* Molecular evolutionary characterization of the mussel *Mytilus* histone multigene family: First record of a tandemly repeated unit of five histone genes containing an H1 subtype with 'Orphon' features. *J. Mol. Evol.* **58**, 131–144 (2004).
64. Šíchová, J., Nguyen, P., Dalíková, M. & Marec, F. Chromosomal evolution in tortricid moths: Conserved karyotypes with diverged features. *PLoS One* **8**, e64520. https://doi.org/10.1371/journal.pone.0064520 (2013).
65. Šíchová, J. *et al.* Dynamic karyotype evolution and unique sex determination systems in *Leptidea* wood white butterflies. *BMC Evol. Biol.* **15**, 89. https://doi.org/10.1186/s12862-015-0375-4 (2015).
66. Mutanen, M., Wahlberg, N. & Kaila, L. Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. *Proc. R. Soc. B* **277**, 2839–2848 (2010).
67. Wiemers, M., Chazot, N., Wheat, C., Schweiger, O. & Wahlberg, N. A complete time-calibrated multi-gene phylogeny of the European butterflies. *ZooKeys* **938**, 97–124 (2020).
68. Kawahara, A. Y. *et al.* Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *PNAS* **116**, 22657–22663 (2019).
69. Marec, F. & Traut, W. Synaptonemal complexes in female and male meiotic prophase of *Ephestia kuehniella* (Lepidoptera). *Heredity* **71**, 394–404 (1993).
70. Zrzavá, M. *et al.* Sex chromosomes of the iconic moth *Abraxas grossulariata* (Lepidoptera, Geometridae) and its congener *A. sylvata*. *Genes* **9**, 279. https://doi.org/10.3390/genes9060279 (2018).
71. Kiauta, B. & Lankhorst, L. The chromosomes of the caddis-fly, *Glyphotaelius pellucidus* (Retzius, 1783) (Trichoptera: Limnephilidae, Limnephilinae). *Genetica* **40**, 1–6 (1969).
72. Lukhtanov, V. A. Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta). *J. Zool. Syst. Evol. Res.* **38**, 73–79 (2000).
73. Seiler, J. Research on the sex-chromosomes of Psychidae (Lepidoptera). *Biol. Bull.* **36**, 399–404 (1919).
74. Seiler, J. Geschlechtschromosomen-Untersuchungen an Psychiden. *Z. Indukt. Abstamm. Vererbungsl.* **31**, 1–99 (1922).
75. Hejníčková, M. *et al.* Absence of W chromosome in Psychidae moths and implications for the theory of sex chromosome evolution in Lepidoptera. *Genes* **10**, 1016. https://doi.org/10.3390/genes10121016 (2019).
76. Dalíková, M. *et al.* New insights into the evolution of the W chromosome in Lepidoptera. *J. Hered.* **108**, 709–719 (2017).
77. Prins, J. D., Prins, W. D. & Dall'Asta, U. The karyotype of *Cameraria ohridella* (Lepidoptera: Gracillariidae). *Phegea* **301**, 5–10 (2002).
78. Kawazoé, A. The chromosome in the primitive or microlepidopterous moth-groups II. *Proc. Jpn. Acad. Ser. B* **63**, 87–90 (1987).
79. Nilsson, N.-O., Löfstedt, C. & Dävring, L. Unusual sex chromosome inheritance in six species of small ermine moths (Yponomeuta, Yponomeutidae, Lepidoptera). *Hereditas* **108**, 259–265 (1988).
80. Ortiz, E. & Templado, J. Los cromosomas de tres especies de tortrícidos (Lep. Tortricidae). *EOS Rev. Esp. Entomol.* **51**, 77–84 (1976).
81. Robinson, R. *Lepidoptera Genetics*. (Pergamon Press, 1971).
82. Carabajal Paladino, L. Z. *et al.* The effect of X-rays on cytological traits of *Tuta absoluta* (Lepidoptera: Gelechiidae). *Fla. Entomol.* **99**, 43–53 (2016).
83. Schulz, H.-J. & Traut, W. The pachytene complement of the wildtype and a chromosome mutant strain of the flour moth, *Ephestia kuehniella* (Lepidoptera). *Genetica* **50**, 61–66 (1979).
84. Van't Hof, A. E. *et al.* Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): A model of industrial melanism. *Heredity* **110**, 283–295 (2013).
85. Kawamura, N. Cytological studies on the mosaic silkworms induced by low temperature treatment. *Chromosoma* **74**, 179–188 (1979).
86. Bytinski-Salz, H. Untersuchungen an Lepidopteren-hybriden V. Die Verwandfschaft der *Platysamia*-Arten (Lepidoptera, Saturniidae) nach Untersuchungen über die Fertilität und die Chromosomenverhältnisse ihrer Bastarde. *Arch. Exp. Zellforsch* **22**, 217–237 (1938).
87. Ahola, V. *et al.* The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat. Commun.* **5**, 4737. https://doi.org/10.1038/ncomms5737 (2014).
88. da Silva, M., Ribeiro, E. D. & Matoso, D. A. Chromosomal polymorphism in two species of *Hypancistrus* (Siluriformes: Loricariidae): An integrative approach for understanding their biodiversity. *Genetica* **142**, 127–139 (2014).
89. Supiwong, W. *et al.* Karyotype diversity and evolutionary trends in the Asian swamp eel *Monopterus albus* (Synbranchiformes, Synbranchidae): A case of chromosomal speciation?. *BMC Evol. Biol.* **19**, 73 (2019).
90. Šíchová, J. *et al.* Fissions, fusions, and translocations shaped the karyotype and multiple sex chromosome constitution of the northeast-Asian wood white butterfly, *Leptidea amurensis*. *Biol. J. Linn. Soc.* **118**, 457–471 (2016).

91. Lukhtanov, V. A., Dincă, V., Talavera, G. & Vila, R. Unprecedented within-species chromosome number cline in the Wood White butterfly *Leptidea sinapis* and its significance for karyotype evolution and speciation. *BMC Evol. Biol.* **11**, 109. https://doi.org/10.1186/1471-2148-11-109 (2011).
92. Hill, J. *et al.* Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Sci. Adv.* **5**, eaau3648. https://doi.org/10.1126/sciadv.aau3648 (2019).
93. Ferguson, L. *et al.* Ancient expansion of the Hox cluster in Lepidoptera generated four Homeobox genes implicated in extra-embryonic tissue formation. *PLoS Genet.* **10**, e1004698. https://doi.org/10.1371/journal.pgen.1004698 (2014).
94. Thomas, G. W. C. *et al.* Gene content evolution in the arthropods. *Genome Biol.* **21**, 15. https://doi.org/10.1186/s13059-019-1925-7 (2020).
95. Cheng, R.-L., Yu, Y.-X., Liu, L.-X., Zhang, C.-X. & Fang, C.-X. A draft genome of the ghost moth, *Thitarodes* (*Hepialus*) sp., a medicinal caterpillar fungus. *Insect Sci.* **23**, 326–329 (2016).
96. Fraïsse, C., Picard, M. A. L. & Vicoso, B. The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nat. Commun.* **8**, 1486. https://doi.org/10.1038/s41467-017-01663-5 (2017).
97. Aguilera, P. M., Debat, H. J., Scaldaferro, M. A., Martí, D. A. & Grabiele, M. FISH-mapping of the 5S rDNA locus in chili peppers (*Capsicum*-Solanaceae). *An. Acad. Bras. Ciênc.* **88**, 117–125 (2016).
98. Volkov, R. A. *et al.* Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant. Biol.* **17**, 21. https://doi.org/10.1186/s12870-017-0978-6 (2017).
99. Mount, S. M., Gotea, V., Lin, C.-F., Hernandez, K. & Makalowski, W. Spliceosomal small nuclear RNA genes in 11 insect genomes. *RNA* **13**, 5–14 (2006).
100. Marz, M., Kirsten, T. & Stadler, P. F. Evolution of spliceosomal snRNA genes in metazoan animals. *J. Mol. Evol.* **67**, 594–607 (2008).
101. Mesa, A., Somarelli, J. A. & Herrera, R. J. Small nuclear RNA variants of three *Bombyx mori* strains. *Entomol. Res.* **38**, 61–68 (2008).
102. Merlo, M. A., Cross, I., Rodríguez-Rúa, A., Manchado, M. & Rebordinos, L. First approach to studying the genetics of the meagre (*Argyrosomus regius*; Asso, 1801) using three multigene families. *Aquac. Res.* **44**, 974–984 (2013).
103. Mediouni, J., Fuková, I., Frydrychová, R., Dhouibi, M. H. & Marec, F. Karyotype, sex chromatin and sex chromosome differentiation in the carob moth, *Ectomyelois ceratoniae* (Lepidoptera: Pyralidae). *Caryologia* **57**, 184–194 (2004).
104. Lockwood, A. "Ringer" solutions and some notes on the physiological basis of their ionic composition. *Comp. Biol. Phys.* **2**, 241–289 (1961).
105. Rovatsos, M., Altmanová, M., Pokorná, M. & Kratochvíl, L. Conserved sex chromosomes across adaptively radiated *Anolis* lizards. *Evolution* **68**, 2079–2085 (2014).
106. Blin, N. & Stafford, D. W. A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.* **3**, 2303–2308 (1976).
107. Ferguson, K. B. *et al.* Jekyll or Hyde? The genome (and more) of *Nesidiocoris tenuis*, a zoophytophagous predatory bug that is both a biological control agent and a pest. *Insect Mol. Biol.* **30**, 188–209 (2020).
108. Fuková, I. *et al.* Probing the W chromosome of the codling moth, *Cydia pomonella*, with sequences from microdissected sex chromatin. *Chromosoma* **116**, 135–145 (2007).

Acknowledgements

We thank Marie Korchová for technical assistance. We further thank David Doležel, Peter Roessingh, Ivo Šauman, Hana Sehadová, and Atsuo Yoshido for material and Jan Provazník for help with data visualization.

Author contributions

P.N. conceived the study. P.N., M.D., and F.M. designed experiments. I.P., M.H., S.V., A.V., L.Z.C.P., and M.Z. performed research. L.Z.C.P. contributed an optimized protocol of TSA-FISH. I.P., M.D., S.V., A.V., M.Z., and P.N. analysed data. I.P. created the figures. I.P. and P.N. wrote the manuscript. All authors read and approved the manuscript.

Funding

This research was funded by the Grant 17-17211S of the Czech Science Foundation (CSF) to PN. Support from CSF Grants 17-13713S and 20-13784S given to FM is also acknowledged.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-91665-7.

Correspondence and requests for materials should be addressed to P.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.2. Chapter II

The role of repetitive DNA in re-patterning of major rDNA clusters in Lepidoptera.

Dalíková M.*, Provazníková I.*, Provazník J., Grof-Tisza P., Pepi A., Nguyen P.
bioRxiv 2022.03.26.485928; doi: 10.1101/2022.03.26.485928 *(under revision in Genome Biology and Evolution)*

* equal contribution

Abstract

Genes for major ribosomal RNAs (rDNA) are present in multiple copies organized in tandem arrays. Number and position of rDNA loci can change dynamically and their re-patterning is presumably driven by repetitive sequences. We explored a peculiar rDNA organization in several representatives of Lepidoptera with either extremely large or numerous rDNA clusters. We combined molecular cytogenetics with analyses of second and third generation sequencing data to show that rDNA spreads as a transcription unit and reveal association between rDNA and various repeats. Furthermore, we performed comparative long read analyses between the species with derived rDNA distribution and moths with a single rDNA locus, which is considered ancestral. Our results suggest that satellite arrays, rather than mobile elements, facilitate homology-mediated spread of rDNA via either integration of extrachromosomal rDNA circles or ectopic recombination. The latter arguably better explains preferential spread of rDNA into terminal regions of lepidopteran chromosomes as efficiency of ectopic recombination depends on proximity of homologous sequences to telomeres.

The role of repetitive DNA in re-patterning of major rDNA clusters in Lepidoptera

Martina Dalíková [1,2,$], Irena Provazníková [1,2,3,$], Jan Provazník [1,3], Patrick Grof-Tisza[4], Adam Pepi[5], and Petr Nguyen [1,2 *]

Affiliation:

1 Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

2 Institute of Entomology, Biology Centre CAS, České Budějovice, Czech Republic

3 Current address: European Molecular Biology Laboratory, Heidelberg, Germany

4 Institute of Biology, Laboratory of Evolutionary Entomology, University of Neuchâtel, Neuchâtel, Switzerland

5 Department of Biology, Tufts University, Medford MA USA

$ These authors contributed equally to this work

*** Correspondence to:**

Petr Nguyen

petr.nguyen@prf.jcu.cz

Key words: FISH, major rDNA, R elements, retrotransposons, transposons, satellites, repetitive DNA, Lepidoptera

Abstract

Genes for major ribosomal RNAs (rDNA) are present in multiple copies organized in tandem arrays. Number and position of rDNA loci can change dynamically and their re-patterning is presumably driven by repetitive sequences. We explored a peculiar rDNA organization in several representatives of Lepidoptera with either extremely large or numerous rDNA clusters. We combined molecular cytogenetics with analyses of second and third generation sequencing data to show that rDNA spreads as a transcription unit and reveal association between rDNA and various repeats. Furthermore, we performed comparative long read analyses between the species with derived rDNA distribution and moths with a single rDNA locus, which is considered ancestral. Our results suggest that satellite arrays, rather than mobile elements, facilitate homology-mediated spread of rDNA via either integration of extrachromosomal rDNA circles or ectopic recombination. The latter arguably better explains

preferential spread of rDNA into terminal regions of lepidopteran chromosomes as efficiency of ectopic recombination depends on proximity of homologous sequences to telomeres.

Introduction

Ribosomal RNAs have a central role in ribosome functions in protein synthesis and thus are a cornerstone for life as we know it (Noller et al., 2017). They are shared by all eukaryotes and have been considered the oldest repetitive fraction (Symonová, 2019) as their genes are present in multiple copies organized in tandem arrays. The genes for major ribosomal RNAs (rDNA), i.e. 18S, 5.8S, and 28S, form a transcription unit, in which internal transcribed spacers (ITS 1 and 2) separate individual genes. In eukaryotic genomes (Prokopowich et al., 2003), there are hundreds or even thousands of rDNA units separated by intergenic spacers (IGS) (Long and Dawid, 1980).

Sequences of rRNA genes and their transcribed spacers have been used in taxonomy for species identification (Wu et al., 2015) or to reconstruct phylogenetic relationships (Fiore-Donno et al., 2012). Moreover, thanks to their cluster organization, the rDNA can be easily detected on chromosomes by fluorescent *in situ* hybridization (FISH), which makes it an important marker in cytogenetic studies (Ferretti et al., 2019; Nguyen et al., 2010; Palacios-Gimenez et al., 2013; Provazníková et al., 2021). The rDNA clusters can be present on autosomes, sex chromosomes or even supernumerary chromosomes, i.e. B chromosomes (Cabral-de-Mello et al., 2011; Poletto et al., 2010; Provazníková et al., 2021; Silva et al., 2014; Zrzavá et al., 2018). While most animal species have only one rDNA locus, up to tens of loci were reported in some extreme cases (Eickbush and Eickbush, 2007; Sochorová et al., 2018 and references therein). The active loci are also called the nucleolar organizer regions (NORs) (Ingle et al., 1975; Kobayashi, 2011), as transcription of major rDNA genes and processing of primary transcripts give rise to a sub-nuclear compartment known as a nucleolus (reviewed in Eickbush and Eickbush, 2007). In general, changes in distribution of rDNA genes are dynamic and rDNA was thus compared to mobile elements (MEs), which, in turn, have been considered an important driver in rDNA re-patterning (Cabral-de-Mello et al., 2011; de Sene et al., 2015; Elliott et al., 2013; Ferretti et al., 2019; Scacchetti et al., 2012).

The order Lepidoptera with its 160,000 species of moths and butterflies represents one of the largest insect radiations (Van Nieukerken et al., 2011). Their rich species and ecological diversity contrast with their conserved genome architecture with the ancestral and the most common chromosome number being n=31 (Ahola et al., 2014; Robinson, 1971; Van't Hof et al., 2013). Detailed analyses of advanced ditrysian species, such as the peppered moth (*Biston* betularia, n=31; Van't Hof et al., 2013), the Glanville fritillary (*Melitaea cinxia*, n=31; Ahola et al., 2014), and the tobacco cutworm (*Spodoptera litura*, n=31; Cheng et al., 2017) showed highly conserved synteny and order of genes between homoeologous chromosomes. A typical lepidopteran mitotic complement consists of small dot-shape chromosomes (Fuková et al., 2005; Mediouni et al., 2004; Prins and Saitoh, 2003), which lack localized centromere, i.e. they are holokinetic (Wolf et al., 1997). Moreover, traditional bending techniques failed to differentiate individual chromosomes, which has made the classic cytogenetic research in Lepidoptera rather challenging (Bedo, 1984) and limited it, for long time, only to chromosome counting (Lukhtanov, 2015; Robinson, 1971). However, the use of various FISH modifications provided great insight into evolution of lepidopteran karyotypes (Van't Hof et al., 2013; Yasukochi et al., 2011), sex chromosomes (Carabajal Paladino et al., 2019; Martina Dalíková et al., 2017a; Šíchová et al., 2015; Vítková et al., 2007), repetitive sequences (Šíchová et al., 2015, 2013), and gene families such as major rDNA (Nguyen et al., 2010; Provazníková et al., 2021).

Number and localization of rDNA loci were determined using FISH with the 18S rRNA probe in various species sampled across Lepidoptera (Fuková et al., 2005; Nguyen et al., 2010; Provazníková et al., 2021; Šíchová et al., 2016, 2015, 2013; Vershinina et al., 2015; Zrzavá et al., 2018). The results implied that one terminal rDNA cluster is probably the ancestral state as it was found across all Lepidoptera. In some ditrysian families, such as Noctuidae and Erebidae, the rDNA cluster moved to interstitial position, which was conserved in all studied species. When multiple clusters are present, they are located terminally in majority of species. Higher numbers of rDNA clusters were observed in representatives of the families Psychidae (3-4 clusters) and Nymphalidae (7-11 clusters), *Biston betularia* (3 clusters, Geometridae) and *Hyalophora cecropia* (3 clusters, Saturniidae), all having the ancestral haploid chromosome number n=31. Thus, spread of rDNA clusters is not clearly associated with large scale chromosome rearrangements such as chromosome fissions or fusions. Unusual distribution

of rDNA was observed in the ghost moth, *Hepialus humuli* (Hepialidae), and the horse chestnut leaf miner, *Cameraria ohridella* (Gracillariidae), in which signal of the 18S rDNA probe covered about one half and one fourth of a single NOR-bearing chromosome, respectively (Provazníková et al., 2021). It was proposed that the dynamic rDNA repatterning is due to ectopic recombination, i.e. recombination between non-homologous regions mediated by ubiquitous repetitive sequences (Nguyen et al., 2010). However, the hypothesis is yet to be tested.

In this study, we decided to explore a peculiar rDNA organization in *H. humuli* and *C. ohridella*, with extremely large rDNA clusters and nymphalids *Aglais urticae* and *Inachis io* with seven and eleven loci per haploid genome, respectively. We performed FISH with probes for 18S and 28S rDNA to test whether genes for major rRNAs spread individually or as a transcription unit. Further, we sequenced genomes of all four species and analysed repetitive sequences and their co-localization with rDNA using the RepeatExplorer pipeline (Novák et al., 2013, 2010). We estimated portion of rDNA units associated with identified repeats by analyses of coverage. The co-localization of several repetitive sequences with rDNA was verified by FISH and in *H. humuli* and the nymphalids also by analysis of long reads. The long reads analysis was further performed also in *Phymatopus californicus* (Hepialidae), to compare it with the peculiar *H. humuli* rDNA organization, and in *Lymantria dispar* (Erebidae), *Spodoptera frugiperda* (Noctuidae), and *Plutella xylostella* (Plutellidae) to compare rDNA composition between species with an ancestral and highly derived rDNA distribution. Our work shows that combining molecular cytogenetic techniques with next generation sequencing technologies represent a powerful tool to study evolution of genome architecture in Lepidoptera.

Results

FISH with 18S and 28S rDNA probes

To examine the organization of the rDNA clusters in genomes of four studied species, namely: *H. humuli*, *C. ohridella*, *A. urticae*, *and I. io*, FISH with 18S and 28S rDNA gene probes was carried out. Hybridization patterns of 18S rDNA probe of all four species correspond to previous results (Nguyen et al., 2010; Provazníková et al., 2021). Moreover, 28S rDNA probe colocalized with 18S rDNA probe in all cases which suggests that the observed patterns of

rDNA distribution are due to spread of the whole rDNA unit. One large major rDNA cluster covering large portion of one chromosomal bivalent was observed in pachytene nuclei of *H. humuli* (Figure 1a) and *C. ohridella* (Figure 1b). Additionally, a strong DAPI-positive heterochromatin block colocalized with major rDNA cluster in *H. humuli* (Figure 1a detail). In pachytene nuclei of *A. urticae* and *I. io*, multiple rDNA clusters were observed as expected. Seven small terminal clusters in pachytene nucleus of *A. urticae* did not colocalize with any heterochromatin blocks (Figure 1c), whereas in pachytene nucleus of *I. io*, 11 terminal signals of various sizes were detected and 6 of them colocalizing with small DAPI-positive blocks (example in Figure 1d detail). Numerous chromosomal bivalents bearing small terminal heterochromatin blocks seem to be typical feature for *I. io* karyotype.

Repeat Explorer analysis

To identify repeats associated with 45S rDNA we performed Repeat Explorer (RE) analysis in all four studied species. Repetitive sequences with frequent colocalization in the genome can be identified through RE analysis as clusters connected by pair-end reads forming the so-called superclusters. In *C. ohridella*, major rDNA genes were split into two clusters. Surprisingly, these clusters have no connection neither between each other nor to any other identified repeat. The estimation of genome proportion formed by major rDNA in this species is about 0.04 % (Suppl. Table S1). In *I. io*, clusters annotated as 45S rDNA were part of supercluster 11. This supercluster was formed by 3 clusters which were annotated as 28S and LINE R2 element (cluster 27), 18S and 5.8S (cluster 35), and putative satellite (IiSat, cluster 37) with predicted monomer length 157 bp (Suppl. Table S1). As cluster 27 was formed by both 28S rRNA gene and LINE R2 elements (IiR2), genome proportion of major rDNA cannot be determined with certainty from the RE results alone; but these genes could comprise 0.12-0.29% of *I. io* genome (Suppl. Table S1). In *A. urticae*, genes for major rDNA were also divided into several clusters, three clusters 17, 21 and 29 annotated as 45S rDNA formed one supercluster 8 and were not connected to any other repeat by 10 or more shared pair-end reads (Suppl. Table S1). However, after further inspection one contig corresponding to cluster 29 contained tandemly repeated sequence suggesting that a satellite repeat (AuSat) with monomer approx. 400 bp is part of this cluster. Based on RE estimate, the major rDNA clusters formed 0.59 % of *A. urticae* genome. In *H. humuli*, major rDNA genes represented 0.17% of the genome and were all comprised in the cluster 53 which was part of the supercluster 24

together with the cluster 67 annotated as ME from Ty3/Gypsy group (Hh Ty3/gypsyA; Suppl. Table S1).

FISH with 18S rDNA and ME probes

To verify the results obtained from RE analysis, we mapped ME probe together with 18S probe on chromosome preparations of *H. humuli* and *I. io* by double TSA FISH. In *H. humuli*, both the 18S rDNA probe and the pooled Hh Ty3/GypsyA *RT*, *INT*, and *RH* probes hybridized to the major rDNA cluster region (Figure 2a) and thus confirmed association between Hh Ty3/GypsyA and the major rDNA. Similar pattern was observed in *I. io*, in which both the 18S rDNA probe and the probe for IiR2 *RT* hybridized to all 11 clusters of major rDNA in pachytene nuclei (Figure 2b), thus proving an association between rDNA and the IiR2 ME. Hybridization of the IiSat probe did not provide any clear signal which was probably caused by low quality of generated probe. Additionally, to investigate whether IiR2 is also present in genome of closely related *A. urticae*, we hybridized the 18S rDNA probe and the IiR2 probe to pachytene nuclei of *A. urticae*. However, no clear hybridization signal was observed (results not shown).

Long read analysis

Long read analysis was used to further verify connection between rDNA and repeats revealed by RE and FISH results. Output of the *H. humuli* sequencing run was poor both in overall yield and read length. After default quality filtering, which was part of the base calling process we obtained 4 Gb in reads with N50 length 6 kb. Due to low coverage of obtained Nanopore data we were able to analyse only 567 reads longer than 15 kb with mean quality (Q) > 10 containing major rRNA genes in *H. humuli*. Most of these reads contained non-functional short copies of three Ty3/Gypsy elements, two LINE elements (from L2 and RTE-RTE groups), two PIF elements (Harbinger and Spy group), and a P element in the IGSs. The organisation and length of the IGSs were highly conserved. Only 18 reads bearing rDNA did not contain any of the mentioned MEs and around 12 % of the reads exhibited some irregularities in the observed pattern (Suppl. Figure S1). All reads containing rDNA were used to assemble rDNA unit including IGS. The resulting Flye assembly contained single circular contig 77 920 bp long corresponding to two complete rDNA units with their lengths differing only in 6bp. Although only one of the observed Ty3/Gypsy elements, Hh Ty3/gypsyA, was detected by the RE analysis as a part of the supercluster containing rDNA, upon careful examination of RE results,

all the IGS repeats were connected by shared pair-end reads. Yet these did not suffice to bind rDNA and all associated MEs in one supercluster (Suppl. Table S1).

To test for the presence of such complex IGS in other representative of the family Hepialidae, we have analysed also Nanopore reads of *P. californicus*. After basecalling with default quality filtration, we obtained 40 Gb of data with read length N50 of 22 kb. Total of 596 reads bearing major rDNA passed the filtering for length > 15 kb and Q > 10. Yet in this case, the IGSs were much smaller and contained only ca. 800 bp long microsatellite region (Suppl. Figure S2) consisting of sequence complementary to insect telomeric repeat TTAGG and 231 bp long region of TTATG microsatellite. The Flye assembly yielded a single 30 616 bp long circular contig. However, this contig corresponded to three complete major rDNA units which differed in length of microsatellite region by up to four TTAGG repeats. The major rDNA unit in *P. californicus* is thus about 10 kb long including IGS region.

Due to high coverage of available HiFi PacBio reads of *I. io,* we were able to analyse 3 276 reads containing major rDNA longer than 15 kbp (Suppl. Figure S3). Of these, 2 625 contained at least 200 bp of the satellite recovered by the RE analysis in their IGS. However, the individual major rDNA units differ in length of this satellite array (Suppl. Table S2). In 946 reads, the R2 element was inserted in major rDNA genes. Surprisingly some of these insertions were not limited to the 28S rRNA gene, suggesting ongoing degeneration of rDNA units via R2 insertions in *I. io*. The attempt to assemble the most prevalent variant of complete rDNA unit in *I. io* failed as all assemblies contained more than 300 contigs of variable length, both linear and circular. However, most of the contigs had very low coverage. Moreover, only 27 contigs contained more than 200 bp of any rRNA gene. Out of all obtained contigs, only three had mean coverage over 10% of used PacBio reads and their length varied from aprox. 15.8-33 kb. These three contigs all contained at least some of the major rRNA genes either with or without R2 element insertion and the satellite array with variable length from 4.5 kb to 17 kb.This further emphasizes the variability in IGSs in this species.

In *A. urticae*, the available HiFi PacBio data contained 2 921 reads bearing major rDNA > 15 kb. The long read analysis confirmed overall lack of MEs associated with rDNA as (Suppl. Figure S4) only 179 reads contained either LINE or Ty3/Gypsy element adjacent to major rRNA genes, including R2 element inserted into 28S rDNA. Despite the absence of MEs in IGS of *A. urticae*, this region seems to vary in length between major rDNA units (Suppl. Figure S4). The

variation of IGS was reflected also in assembly results as any assembly produced by Flye contained over 70 linear contigs in the length ranging 19 – 69 kb. However, only six contigs contained more than 200 bp of major rDNA and only one of these contigs had mean coverage over 10% of input reads. This contig is 31 kbp long and contains two complete rDNA units including IGS regions. One unit contains R2 insertion in 28S gene and based on dot plot both IGSs contain approx. 2kb of satellite (AuSat) array. Both AuSat IGS satellite arrays contain 6 monomers with variable length from 252 to 408 bp with the last monomer being the shortest one. This satellite was present in most major rDNA units as it was found in 2805 reads containing major rDNA (Suppl. Figure S4). After further inspection of previous RE results in this species, a sequence homologous to the satellite was found among contigs belonging to the cluster 29 (Suppl. Table S1) which also contained a part of the 28S rRNA gene. Interestingly, our RE results did not contain any cluster with sequence homologues to the AuR2 element found in long reads. Considering that samples for RE were sampled from the Czech *A. urticae* population while specimen sequenced by PacBio originated from Great Britain, the R2 insertion into rDNA units may represent inter-population variation in this species.

To test if the variable and/or long IGSs are connected with atypical rDNA genomic organization we performed long read analysis also in species with one major rDNA locus per haploid genome which is supposedly ancestral in Lepidoptera (Nguyen et al., 2010; Provazníková et al., 2021), namely in *Plutella xylostella*, *Spodoptera frugiperda* and *Limatria dispar*. In *P. xylostella*, we analysed 419 PacBio reads containing major rDNA with length at least 15 kb. The rDNA was not associated with any ME, however the IGS contained 850 bp satellite (PxSat) region (Suppl. Figure S5). This array consisted of four monomers with slightly variable length between 248-258 bp with the last monomer being incomplete and only 91 bp long. At least 200 bp of this satellite array was found in 343 reads out of the analysed reads containing major rDNA. The Flye assembly of all filtered rDNA bearing PacBio reads yielded one circular contig approximately 21 kb long. This contig consists of two similarly long complete rDNA units (10.7 and 10.9 kb) which differ in the PxSat array length by 242 bp.

In *L. dispar*, 242 quality filtered PacBio reads contained major rDNA. Surprisingly, major rDNA in this species was associated with two different MEs from LINE R1 group specific for rDNA (Suppl. Figure S6). 51 analysed reads contained at least one of those elements and

8 reads contained both transposons. Flye assembly contained 2 linear contigs 11 kbp and 9.4 kb long with the latter having approx. 10x higher coverage. Neither of these contigs contained complete major rDNA unit, the longer contig contained both R1 elements and the shorter one incomplete major rDNA unit with IGS but only partial 28S rRNA gene.

In *S. frugiperda*, we analysed 115 quality and length filtered PacBio reads containing major rDNA. There were no MEs or satellite sequences observed in these reads (Suppl. Figure S7). Flye assembly consisted of only one circular contig 17.9 kb long which contained two identical complete major rDNA units.

Satellite DNA arrays contained in IGSs of *P. xylostella* representing the ancestral rDNA distribution and both nymphalid species with multiple rDNA clusters seemed to vary in length. Thus, we further characterised these satellite arrays. All three species have similar most represented array length in the PacBio reads, as the medians are ranging from 1.74 to 2.21 kb (Suppl. Table S2, Suppl. Figure S8). However, they differ greatly in maximal observed length, which was 4.67 kb in *P. xylostella* but over 15kb in both nymphalid species (Suppl. Table S2, Suppl. Figure S8). Similar differences can be seen in the satellite array length in the rDNA assemblies. In nymphalid species we obtained multiple contigs with variable satellite length ranging from less than 1 kb to 2.23 kb in *A. urticae* and over 17 kb in *I. io*. (Suppl. TableS2). Whereas, in *P. xylostella* we obtained just one contig with two satellite arrays differing inlength by just 242 bp (see above). These results suggest higher variation in length of IGS satellite arrays in *A. urticae* and *I. io* compared to *P. xylostella* (Suppl. Table S2, Suppl. Figure S8). Moreover, the three species differed in the presence of PacBio reads containing satellite sequence without any part the rDNA unit. While we found no such reads in *P. xylostella*, 11 and 30 reads were found in *A. urticae* and *I. io* PacBio data, respectively. As the lengths of PacBio reads bearing only satellite in both nymphalid species exceed the lengths of observed satellite arrays in rDNA assemblies (Suppl. Table S2), these reads either come from very large IGS satellite arrays or they may represent satellite arrays outside the rDNA cluster. The latter is supported by the recently published *A. urticae* genome (Bishop et al., 2021), which contains only 6 terminal rDNA clusters, however, AuSat is found both within the IGS region and right outside two NORs (Suppl. Figure S9).

Coverage analysis

Paired-end reads produced by Illumina sequencing provided us with sufficient coverage to compare per base abundance of reads aligned to the consensus sequences of the whole rDNA unit of *H. humuli* obtained via long reads analysis (see above) and the most represented complete rDNA unit of *I. io* and *A. urticae* from rDNA assembled contigs (see above). In case of *H. humuli* (Figure 3a, Suppl. Table S3), the repetitive elements associated with major rDNA are most likely present elsewhere in the genome as we observed a uniform coverage of the rDNA genes and varying but higher coverage of the intergenic MEs. In *I. io*, we observed that the R2 element is only present in roughly one third of the copies of the 28S rRNA gene (Figure 3b, Suppl. Tab. S3). The coverage of IiSat region was approx. two times larger compared to rRNA genes (Figure 3b, Suppl. Tab. S3), which similarly to the results obtained from PacBio reads suggest the IiSat presence outside of rDNA clusters and/or the variable length of this satellite array inside IGS regions. Surprisingly, in *A. urticae* all the rDNA unit elements showed even coverage including the AuSat region (Figure 3c, Suppl. Tab. S3). This discrepancy between results obtained through illumina and PacBio reads may represent another inter- populational variability in this species in the repeat content (see above).

Discussion

The arrays of major rRNA genes have become a very popular cytogenetic marker in comparative studies of karyotype evolution. Distribution of major rDNA can be relatively stable or rather dynamic in various taxa (Cabral-de-Mello et al., 2011; García-Souto et al., 2016; Nguyen et al., 2010; Perumal et al., 2017) and even intraspecific variability was observed (Baumgärtner et al., 2014; Ferretti et al., 2019; Šíchová et al., 2015). Changes in number and localization of rDNA loci have been ascribed to sequence homogeneity maintained by gene conversion (reviewed in Eickbush and Eickbush, 2007) and chromosomal rearrangements mediated by ectopic recombination (Cabral-de-Mello et al., 2011; Ferretti et al., 2019; Nguyen et al., 2010), transposition (Raskina et al., 2008, 2004) or integration of extrachromosomal circular rDNA (ecc-rDNA; Proux-Wéra et al., 2013).

To gain some insight into mechanism of spread of major rDNA in genomes of moths and butterflies, we examined four lepidopteran species with peculiar distributions of 18S rDNA (Nguyen et al., 2010; Provazníková et al., 2021) namely *H. humuli* and *C. ohridella* with large rDNA locus covering up to half of the chromosome, and *A. urticae* and *I. io*, in which multiplication of rDNA loci occurred, and compared them to three control species, namely *P. xylostella*, *S. frugiperda* and *L. dispar*, with a single major rDNA locus. In addition, *H. humuli* was compared to another hepialid, *P. californicus*. Combining molecular cytogenetics and sequencing of both second and third generation, we explored co-localization of 18S and 28S rDNA, association of major rDNA units with repetitive sequences, and their potential influence on evolution of major rDNA.

To determine the position and number of rDNA clusters, only one probe, often 18S rDNA, is hybridized on chromosomes as a representative of the whole major rDNA unit. It is usually assumed that major rDNA units spread as a whole (Bueno et al., 2013). However, Ferretti et al., (2019) discovered high intraspecific and interspecific variability and independent mobility of each component of the major rDNA unit (18S, ITS1, 5.8S, ITS2, and 28S) in the genome of six different populations of a grasshopper *Abracris flavolineata*. Similar pattern was also observed in *Coregonus* fishes where both complete and partial rDNA units were detected by FISH mapping of individual rRNA genes (Symonová et al., 2013). Therefore, we physically mapped partial sequences of 18S and 28S rDNA to test whether rDNA spreads as a whole unit in the studied species. In *H. humuli* and *C. ohridella*, both 18S and 28S rDNA

signals co-localized and covered significant portion of one chromosome pair (Figure 1a, b) as reported earlier (Provazníková et al., 2021). Seven and eleven rDNA loci were highlighted by both probes in *A. urticae* and *I. io*, respectively (Figure 1c, d), which is in agreement with known distribution of their major rDNA (Nguyen et al., 2010; Provazníková et al., 2021). Thus, it is reasonable to conclude that a complete major rDNA unit was amplified or spread into new loci in the species under study. This is further corroborated by coverage analyses carried out in *H. humuli*, *A. urticae* and *I. io*, which showed similar read depth across their rDNA units (Figure 3 a, b, c, Suppl. Table S3).

For its dynamic evolution, rDNA has been often compared to repetitive sequences as arrays of rDNA are often found within heterochromatin (Lohe and Roberts, 2000). Fragments of rDNA can be amplified into satellite-like tandem arrays (Lohe and Roberts, 2000) and were found to be associated with satellites and other repeats (Barbosa et al., 2015; Jakubczak et al., 1991; Raskina et al., 2008; Sember et al., 2018; Symonová et al., 2013) which could mediate their spread (cf. Raskina et al., 2008; Nguyen et al., 2010; Proux-Wéra et al., 2013). Therefore, we clustered paired-end illumina reads of species under study using the RE pipeline (Novák et al., 2010) and searched for association of identified repeats with major rRNA genes. The paired-end reads did not link major rDNA with any other clustered repeats in *C. ohridella* and *A. urticae*, although we could not have excluded that such repeats are present either in low frequencies or distance bigger than the library insert size (see below). Yet, the association was recovered between major rDNA and Ty3/gypsy retrotransposon in *H. humuli* (Hh Ty3/gypsyA), and R2 element (IiR2) and a satellite (IiSat) in *I. io* (Suppl. Table S1).

Third generation sequencing technologies have recently provided and unprecedented insight into organisation of repetitive sequences including rDNA genes (e.g. Belser et al., 2021; Sims et al., 2021; Sproul et al., 2020; Vondrak et al., 2021). We took advantage of *A. urticae* and *I. io* PacBio data recently released by the Darwin Tree of Life project and analysed long reads, which contained major rDNA. The *I. io* data confirmed our previous results as the *I. io* long reads contained both the IiR2 and IiSat sequences (Suppl. Figure S3, Figure 4). Both proportion of long reads (Suppl. Figure S3) and coverage analysis (Figure 3c, Suppl. Table S3) suggest that IiR2 is present in about one third of rDNA units, which is in agreement with findings from other species (cf. Jakubczak et al., 1991; Zhou et al., 2013). In accordance with the results of RE, we found no MEs associated with rDNA in vast majority of reads in *A. urticae*.

Small fraction of reads, which contained retrotransposon sequences of Ty3/Gypsy, R1 and R2 (Suppl. Figure S4), most likely corresponds to pseudogenes resulting from birth-and-death process (cf. Martí et al., 2021). Yet in contrast to the RE analysis, we found also satellite arrays (AuSat) in the IGS regions in majority of the *A. urticae* long reads bearing major rDNA (Suppl. Figure S4, Figure 4). Surprisingly, different satellites are associated with major rDNA in the two closely related nymphalids. In both species, spacers notably varied in their length which points to a possible lack of concerted evolution. We hypothesize this is due to high number of major rDNA loci, which are not all transcriptionally active. Thus, they do not associate in nucleolus and evolve independently. Alternatively, the observed variation could be ascribed to rDNA subtypes with tissue specific expression or to mutations impairing chromatin modification enzymes (cf. Havlová et al., 2016).

In *H. humuli*, the Hh Ty3/gypsyA retrotransposon was inserted at the very end of rDNA unit, at the junction between 28S rDNA gene and external transcribed spacer (Suppl. Figure 1, Figure 4). Mapping of its partial sequence by TSA FISH revealed its clear co-localization with rDNA (Figure 2a-d). Although we did not detect any other Hh Ty3/gypsyA loci, we cannot exclude it is present elsewhere in the genome as interspersed repeats as combined length of used probes is still under ca. 1300 bp detection limit of the TSA FISH protocol used (cf. (Carabajal Paladino et al., 2014). Indeed, the coverage analysis suggests that abundance of MEs associated with rDNA is higher than abundance of rRNA genes themselves (Figure 3a, Suppl. Table S3). Furthermore, the Hh Ty3/gypsyA copy associated with major rDNA is non-autonomous. It represents only a portion of the corresponding RE cluster, which can be assembled into a complete retrotransposon with LTR repeats. Yet the Hh Ty3/gypsyA sequences in IGS lack long terminal repeats and most of protein coding domains (Suppl. Figure 1, Figure 4). Similar association between major rDNA and the Ty3/gypsy retrotransposon *Beon1 (*Galadriel clade) was observed also in the beet *Beta vulgaris* where, however, the ME is inserted into 18S rRNA genes (Weber et al. 2013).

While short paired-end reads revealed only association between rDNA and Hh Ty3/gypsyA retrotransposon in *H. humuli*, long ONT reads showed that fragments of eight different elements were inserted in IGS (Suppl. Figure 1, Figure 4). Like the Hh Ty3/gypsyA (see above), none of the major rDNA associated repeats is autonomous and thus cannot multiply on their own. However, their transmission is ensured by hitchhiking along with the

indispensable rRNA gene family. It is not clear how expansion of IGS effected transcription of major rRNA genes. The IGS expanded to ca. 39 kb and it is roughly on par with 45 kb long IGS of mouse (Grozdanov et al., 2003) and thus not necessarily detrimental. If the IGS expansion decreases expression of major rRNA genes, increase of copy number would be favoured, which would explain the extraordinary size of the *H. humuli* major rDNA cluster. Moreover, it is possible that rDNA did not actually spread along the chromosome. Rather the total size of the array could have increased as major rDNA unit expanded due to insertion of repeats into IGS. Comparison of rDNA sequences between *H. humuli* and another hepialid *P. californicus,* showed that the expansion of IGS is not shared across the family Hepialidae. *P. californicus* IGS and major rDNA transcription unit have in total only ca. 10 kb. Surprisingly, the IGS contains an array of insect telomeric motif $TTAGG_{(n)}$ and TTATG microsatellite (Suppl. Figure 2, Figure 4; cf. Ruiz-Herrera et al., 2008; Scali et al., 2016; Sember et al., 2018). IGSs are known to contain repetitive motives, which usually do not correspond to MEs (Grozdanov et al., 2003; Havlová et al., 2016). However, association between rDNA and MEs has been reported with 5S being involved more often than 45S rDNA (da Silva et al., 2016; Yano et al., 2020). Insertions of MEs into IGS observed in *H. humuli* thus represents interesting case of complex repeat organization (cf. Vondrak et al., 2021).

To pinpoint a mechanism responsible for changes in distribution of major rDNA in Lepidoptera, we took advantage of available long read sequencing data and compared structure of major rDNA units between species with the highly derived rDNA distribution (see above) with three species with a single rDNA locus, namely *P. xylostella*, *L. dispar*, and *S. frugiperda* (Nguyen et al., 2010; Provazníková et al., 2021). While we found MEs and/or satellite arrays at least in part of the IGS in all species with extraordinary major rDNA patterns but *C. ohridella* for which long reads have not been available, no repetitive sequences were associated with major rDNA in *S. frugiperda* (Suppl. Figure S7, Figure 4). However, satellite arrays (PxSat) of variable size were found in IGS region in *P. xylostella* (Suppl. Figure S5, Figure 4) and analysis of *L. dispar* long reads revealed two types of R1 retrotransposon (Suppl. Figure S6) with 52.3 % nucleotide identity in small fraction of reads.

The R1 and R2 non-LTR retrotransposons are, along with the Pokey DNA transposon, among the few known rDNA-specific elements with insertion sites in the gene for 28S rRNA (Eickbush and Eickbush, 2007, Elliott et al., 2013). The R elements represent one of the oldest

groups of metazoan MEs (Kojima and Fujiwara, 2005). They were described for the first time in a fruit fly, *D. melanogaster*, by Roiha et al., (1981) and afterwards reported in many other animal phyla (reviewed in Eickbush and Eickbush, 2015). Within the order Lepidoptera, the R1 and/or R2 elements have so far been detected only in several species, namely *Bombyx mori* and *B. mandarina* (both Bombycidae), *Manduca sexta* (Sphingidae), four representatives of the family Saturniidae, and *L. dispar* (Erebidae) (Fujiwara et al., 1984; Jakubczak et al., 1991; Kierzek et al., 2009). The latter is of particular interest as our observation of the two R1 retrotransposons and no R2 element in the same species suggests interpopulation differences and rapid turnover of the R1 and R2 retrotransposons. This is further supported by the absence of the R2 element in the Czech *A. urticae* compared to the British population. The R elements were found also in the nymphalids *A. urticae* and *I. io* with seven and eleven rDNA loci, respectively. It was proposed that the R2 retrotransposon plays an important role in maintaining rDNA copy numbers in *Drosophila* (Nelson et al., 2021). Yet, their absence in *S. frugiperda* and *P. xylostella* does not destabilize their rDNA loci and it seems unlikely that the R1 and R2 retrotransposons could mobilize rDNA in species under study. Although it was shown that the R1 and R2 retrotransposons can insert in target site outside 28S rDNA in *B. mori* (Xiong et al., 1988), the IiR2 elements were not detected outside the major rDNA loci in *I. io* (Figure 2b, 3b). Moreover, insertion of the R1 and R2 elements into the 28S rDNA cause pseudogenization of the corresponding rDNA units (Long and Dawid, 1979). The only other MEs associated with major rDNA were observed in *H. humuli.* However, these were not autonomous and their sequence diverged from those found in the rest of the genome. Thus, it seems unlikely that MEs could mediate spread of rDNA observed in some Lepidoptera (Provazníková et al., 2021) via transposition.

On the contrary, satellite arrays such as those found in IGS of nymphalids with high number of rDNA loci, *A. urticae* and *I. io* (Figure 4), could facilitate homology-mediated spread of rDNA via either ectopic recombination or integration of extrachromosomal rDNA circles (Muirhead and Presgraves, 2021; Nguyen et al., 2010; Proux-Wéra et al., 2013; Sproul et al., 2020). Yet, satellite arrays were associated with rDNA also in *P. xylostella* (Figure 4), which has only one rDNA locus. There is a difference between satellite DNA associated with rDNA in *P. xylostella* and the two nymphalids. In *P. xylostella*, we did not find any long reads bearing the PxSat without rDNA or at least partial IGS sequence. In both nymphalids, however, we

identified more than 10 long reads bearing only the satellite arrays, with mean IGS and filtered read lengths being similar in all species which points to a presence of the satellites outside rDNA arrays. Unfortunately, coverage analyses are uninformative for satellites as there is a considerable variation in number of their monomers per rDNA unit (Suppl. Figure S5). Inspection of the *A. urticae* genome assembly suggests that the AuSat is localized only in rDNA clusters and arrays adjacent to rDNA (Suppl. Figure S4), which suggests that rDNA spread into the AuSat loci.

We cannot distinguish with certainty whether the rDNA spread occurred via ectopic recombination or integration of ecc-rDNA. Yet we argue that preferential spread of rDNA into terminal regions of lepidopteran chromosomes (Nguyen et al., 2010; Provazníková et al., 2021) favours ectopic recombination as its efficiency depends on proximity of homologous sequences to telomeres (Goldman and Lichten, 1996; Nguyen et al., 2010), whereas ecc-rDNA could be integrated anywhere in the genome as long as a homologous sequence is present. Little is known about satellite DNA in Lepidoptera, which has been studied in detail only in a dozen of species (Lu et al., 1994; Mandrioli et al., 2003; Mahendran et al., 2006; Věchtová et al., 2016; M. Dalíková et al., 2017b; Cabral-de-Mello et al., 2021). Yet, it seems that abundance of satellite DNA in lepidopteran genomes is very low with scattered distribution and possible enrichment on sex chromosomes (Cabral-de-Mello et al., 2021). This does not reflect distribution of rDNA in Lepidoptera (Nguyen et al., 2010; Provazníková et al., 2021). Yet we cannot exclude those satellites associated with major rDNA are limited to chromosome ends similar to *P. californicus*, which contains telomeric repeats in its IGS (Suppl. Figure S2).

Compared to other regions of rDNA arrays, IGS are rarely studied, and we thus cannot tell whether the observed complex association between rDNA and repetitive sequences (Figure 4) represent a common phenomenon. Our results show that long read sequencing is a valuable tool to study association of repeats including major rDNA as it provided more detailed information about major rDNA associated repeats than analysis of short reads limited by library insert size. Moreover, the long read analysis provided better genomic representation compared to the genome assembly based on these long reads as seen in the *A. urticae* example (Suppl. Figure 4 and 9). Available target enrichment of major rDNA and other repeats for long read sequencing (McKinlay et al., 2021) could provide further insight into formation of complex repeat structures involving rDNA.

Material and methods Material

Specimens of all studied species were collected from wild populations. Females of *Hepialus humuli* were collected in Bochov, Czech Republic and let lay eggs in plastic containers. Hatched larvae were transferred and extensively reared in outdoor pots with planted carrot (*Daucus carota*). Larvae of *Cameraria ohridella* were collected in České Budějovice, Czech Republic, from leaves of the horse-chestnut, *Aesculus hippocastanum*. Specimens of *C. ohridella* were processed immediately after collection. Larvae of two nymphalids, *Inachis io* and *Aglais urticae*, were collected near Vrábče, Czech Republic. They were kept on the common nettle (*Urtica dioica*) in ambient conditions. Larvae of *Phymatopus californicus* were collected from the yellow bush lupine, *Lupinus arboreus*, in the Bodega Marine Reserve (California, USA). Larvae were used for chromosomal preparations and extraction of genomic DNA (gDNA) shortly after collection.

Chromosomal preparation

Chromosomal preparations were prepared by a spreading technique as described in Mediouni et al., (2004) and Dalíková et al., (2017a) with 10 min hypotonization of tissue. Meiotic and mitotic preparations were obtained from gonads of late larval instars of all four species. Afterwards, prepared slides were dehydrated in an ethanol series (70, 80 and 100% ethanol, 30 sec each) and stored at -20 and -80 °C until further use. Remaining tissues were frozen for subsequent gDNA extraction.

Genomic DNA extraction

For downstream applications such as PCR, gDNA was extracted from larvae using NucleoSpin DNA Insect (Macherey-Nagel, Düren, Germany) according to manufacturer's protocol. To obtain high-molecular-weight gDNA for NGS sequencing, gDNA was extracted from larvae using MagAttract HMW DNA Kit (Qiagen, Hilden, Germany) or Nanobind Tissue Big DNA kit (Circulomics Inc, Baltimore, MD, USA) according to manufacturer's protocol. Concentration of extracted samples was measured by Qubit 3.0 Fluorometer (Invitrogen, Carlsbad, CA, USA)and visualized on agarose gel. A single male larva was used as input material for all species but *C. ohridella*, for which 5-10 individuals (larvae and pupae) of both sexes were pooled

because of their small size. A single male adult was used for extraction for Nanopore sequencing.

Repeat Explorer analysis

For analysis of repetitive DNA content, whole gDNA was sequenced on the Illumina platform generating either 150 bp pair-end reads from library with mean insert size 450 bp (Novogene Co., Ltd., Beijing, China) or 250 bp PE reads with the mean insert size 700 bp in case of *C. ohridella* (Genomics Core Facility, EMBL Heidelberg, Germany). The raw reads were quality filtered and trimmed to uniform length of 120 bp (230 bp for *C. ohridella*) by Trimmomatic 3.2 (Bolger et al., 2014). Random sample of two million (one million for *C. ohridella*) trimmed PE reads was analysed by RE pipeline (version cerit-v0.3.1-2706) implemented in Galaxy environment (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/) with automatic annotation via blastn and blastx using the Metazoan 3 Repeat Explorer database. The resulting html files were searched for clusters annotated as major rDNA and their connection to other clusters.

Probes for FISH experiment

All mapped sequences were amplified by PCR using specific primers (for details see in Table 1), purified from agarose gel, and cloned into Promega pGem T-Easy Vector (Promega, Madison, WI, USA). Selected clones were isolated using Nucleo Spin Plasmid kit (Macherey-Nagel) and verified by sequencing (SEQme, Dobříš, Czech Republic).

Fragments of 18S and 28S rRNA genes were amplified from gDNA of codling moth, *Cydia pomonella* (Tortricidae) (cf. Nguyen et al., 2010). To obtain probes, these fragments were re-amplified by PCR from plasmids using specific primers (Table 1), purified by Wizard® SV Gel and PCR Clean-Up System (Promega), and labelled using nick translation protocol by (Kato et al., 2006) with modifications described in (Martina Dalíková et al., 2017a). The 20µL labelling reaction contained 1 µg DNA; 0.5 mM each dATP, dCTP, and dGTP; 0.1 mM dTTP; 20 µM labelled nucleotides; 1× nick translation buffer (50 mM Tris–HCl, pH 7.5; 5 mM MgCl2; 0.005% BSA); 10 mM β-mercaptoethanol; 2.5×10^{-4} U DNase I, and 1 U DNA polymerase I (both ThermoFisher Scientific, Waltham, MA). The reaction was incubated at 15 °C for 45 minutes and enzymes were inactivated at 70 °C for 10 minutes. The 18S rDNA probe was labelled either with biotin-16-dUTP (Roche Diagnostics, Basel, Switzerland) or DNP-11-dUTP

(Jena Bioscience, Jena, Germany) and 28S rDNA probe was labelled by digoxigenin-11-dUTP (Roche Diagnostics) or fluorescein-12-dUTP (Jena Bioscience).

Three sequence fragments of Hh Ty3/GypsyA (CL67 contig 3 and 10) found in *H. humuli* and one of R2 element (CL27contig8) and satellite (CL37 contig4) found in *I. io*, were separately labelled by PCR using plasmid DNA as template according to Provazníková et al., (2021). The 25µL labelling reaction contained 1 – 10 ng template plasmid DNA, 1x Ex Taq buffer, 1 mM each dATP, dCTP, and dGTP; 0.36 mM dTTP; 0.62 mM labelled nucleotides of Fluorescein-12-ddUTP (Jena Biosciences), 5 µmol of each specific primer (**Table 1**), and 0.25 U TaKaRa Ex Taq DNA polymerase (TaKaRa, Otsu, Japan). The resulting labelled probes were purified using Sephadex gel filtration (Illustra Sephadex G-50 fine DNA grade).

Table 1: Primers used for PCR amplification of FISH probe templates and labelling.

Gene	Species	Forward/Reverse primer	T_a	Reference
18S rDNA	*Cydia pomonella*	CGATACCGCGAATGGCTCAATA/ ACAAAGGGCAGGGACGTAATCAAC	58°C	Fuková et al. 2005
28S rDNA	*Cydia pomonella*	GCAGATCTTGGTGGTAGTAGCA/ GATGTACCGCCCCAGTCAAA	58°C	This study
Hh Ty3/GypsyA 1 CL67contig10	*Hepialus humuli*	AAATAAACTCTTAAAAGATGGAGT/ TAATCTCCACTTCTTTTTCCC	58°C	This study
Hh Ty3/GypsyA 2 CL67contig3	*Hepialus humuli*	TTCGATTGAGGGTGATAGGCG/ TCTCAAGCCTATCCAATCGCA	58°C	This study
Hh Ty3/GypsyA 3 CL67contig3	*Hepialus humuli*	TCTTGATCCTGGGTCTTTTACGTT/ CGCGCTATTGGTAGTGTGCT	58°C	This study
Iio R2 CL27contig8	*Inachis io*	CCCAACAGAGAACACCCTCTC/ GTGTTGGGGGATAGCAGGAAA	58°C	This study

FISH with 18S and 28S rDNA probes

Indirect FISH was carried out according to Fuková et al., (2005) and Zrzavá et al., (2018) with some modifications and using two probes, biotin-labelled 18S rDNA probe and digoxigenin-labelled 28S rDNA probe, simultaneously. This technique was used to localize 18S and 28S rDNA genes in genome of *H. humuli*, *C. ohridella*, and *A. urticae*. The slides were dehydrated in an ethanol series (70, 80, and 100% ethanol, 30 sec each) and pre-treated with RNase A (200 ng/µL) in 2× SSC at 37°C for 1 h, washed twice in 2×SSC at RT for 5 min each and incubated in 5×Denhardt's solution at 37°C for 30 minutes. After, slides were denaturated in 70% formamide in 2x SSC at 68°C for 3.5 minutes and immediately dehydrated in an ethanol series (cold 70% for 1 min, 80 and 100% for 30 sec each). Hybridization probe mix containing 10% dextran sulfate, 50% deionized formamide, 25 µg of sonicated salmon sperm and 50 ng of each probe in 2× SSC in final volume of 10 µl was denaturated at 90°C for 5 minutes and immediately placed on ice for 2 minutes. Afterwards, hybridization probe mix was applied on the slide, covered by cover slip, and placed into a humid chamber. Hybridization was carried out at 37°C overnight (12-16 h).

Next day, slides were incubated three times in 50% formamide in 2x SSC and followed by three washes in 2x SSC, both at 46°C for 5 minutes. Slides were then washed three times with 0.1x SSC at 62°C for 5 minutes and once in 1% Triton X in 4x SSC at RT for 10 minutes. The slides were blocked with 2.5% BSA in 4x SSC at RT for 30 minutes, incubated with anti-DIG1 (mouse anti-digoxigenin, 1:100, Roche Diagnostics, Basel, Switzerland) and streptavidin Cy3 conjugate (1:1000, Jackson ImmunoRes. Labs. Inc, West Grove, PA, USA) in 2.5% BSA in 4x SSC at 37°C for 1 hour and washed three times with 1% Triton X in 4x SSC at 37°C for 3 minutes each. To amplify the signals, last three steps were repeated twice, firstly with anti-DIG2 (sheep anti-mouse Ig digoxigenin conjugate, 1:200, Merck Millipore, Billerica, MA, USA) and antistreptavidin (1:25, Vector Labs. Inc, Burlingame, CA, USA) and secondly with anti-DIG3 (sheep anti-digoxigenin fluorescein conjugate, 1:200, Roche Diagnostics) and streptavidin-Cy3 conjugate (1:1000, Jackson ImmunoRes. Labs. Inc). After the last washing step, slides were incubated in 1% Kodak PhotoFlo at RT for 1 minute and counterstained with 0.5 mg/mL DAPI

(4',6-diamidino-2- phenylindole, Sigma-Aldrich, St. Louis, MO, USA) in antifade based on DABCO (1,4-diazabicyclo(2.2.2)-octane; Sigma–Aldrich).

Double TSA FISH

Double FISH with tyramide signal amplification (double TSA FISH) was performed according to Carabajal Paladino et al., (2014) with some modifications. Due to its high sensitivity, double TSA FISH was employed to localize 18S and 28S rDNA genes in *I. io* genome, and ME sequences and 18S rDNA gene in genomes of *I. io* and *H. humuli.* Briefly, slides were dehydrated in an ethanol series (70,80 and 100% ethanol, 30 sec each) and pre-treated with 50 µg/mL pepsin in 0.01 M HCl at 37°C for 10 min, 1% H_2O_2 in 1x PBS at RT for 30 min, and in RNase A (100 µg/mL) in 1x PBS at 37°C for 1 hour. After each pre-treatment, the slides were washed three times in 1x PBS at RT for 5 min each washing. After the last washing, slides were incubated in 5x Denhardt's solution at 37°C for 30 minutes. Directly after the last incubation, 50 µl of hybridization probe mix containing 10% dextran sulphate, 50% deionized formamide, and 10-20 ng of each probe in 2x SSC was applied onto the slide, covered by cover slip, and incubated at 70°C for 5 minutes. Afterwards, slides were placed into the humid chamber and hybridized at 37°C overnight (12-16 h). In *I. io* experiments, the 18S rDNA probe (10-20 ng) labelled with dinitrophenyl (DNP) was used with fluorescein-labelled R2 probe (10-20 ng), or fluorescein-labelled 28S rDNA probe (10-20 ng). In case of *H. humuli*, combination of three fluorescein-labelled ME probes (Hh Ty3/GypsyA 1-3, 5-10 ng each) and 18S rDNA probe (10-20 ng) was used.

The next day, slides were incubated three times 5 min in 50% formamide in 2x SSC at 46°C each, washed three times in 2x SSC at 46°C for 5 minutes each and in 0.1x SSC at 62°C for 5 minutes each, and washed once in 1x TNT at RT for 5 minutes. The slides were blocked in TNB buffer at RT for 30 minutes and incubated with antifluorescein-HRP Conjugate (PerkinElmer) in TNB (diluted 1:1000) at RT for 1 hour. Afterwards, the slides were washed three times in 1x TNT at RT for 5 minutes each and incubated with TSA Plus Fluorescein (PerkinElmer) according to the manual at RT for 3-15 minutes (3-5 min in *H. humuli* and *C. ohridella*, 10-15 min for *I.io*) and washed again three times in 1x TNT at RT for 5 minutes each. To perform the second round of detection and to quench peroxidase activity from previous steps, slides were incubated in 1% H_2O_2 in 1x PBS at RT for 30 min. Next, the slides were washed three times in 1x TNT at RT for 5 minutes each and the amplification steps were

repeated using anti-DNP-HRP Conjugate (PerkinElmer) and TSA Plus Cyanine 3 (PerkinElmer). After the last washing step, the slides were incubated in 1% Kodak PhotoFlo at RT for 1 minute and counterstained with 0.5 mg/mL DAPI (Sigma–Aldrich) in antifade DABCO (Sigma–Aldrich).

Microscoping and image processing

Chromosome preparations from FISH experiments were observed by Zeiss Axioplan 2 microscope (Carl Zeiss, Jena, Germany) equipped with appropriate fluorescence filter sets. An Olympus CCD monochrome camera XM10 equipped with cellSens 1.9 digital imaging software (Olympus Europa Holding, Hamburg, Germany) was used to record and capture black-and-white pictures. Images were captured separately for each fluorescent dye and then pseudocoloured and superimposed with Adobe Photoshop CS4, version 11.0.

Long read sequencing and analysis

High molecular weight DNA from *H. humuli* was enriched for fragments longer than 10 kbp by Short Read Eliminator (Circulomics Inc). The library was prepared by Ligation Sequencing Kit SQK-LSK110 (Oxford Nanopore Technologies, Oxford, UK) according to the manufacture's protocol and therein recommended third party consumables. The library was snap-frozen and stored over night at -70°C and then sequenced using flowcell R10.3 and MinION Mk1B (Oxford Nanopore Technologies). Reads were basecalled by guppy 4.4.1. with high accuracy flip-flop algorithm. The data was filtered for reads 15kbp and longer with quality score over 10 using NanoFilt (De Coster et al., 2018).

Quality and length filtered reads were searched for presence of major rDNA using blastn. Reads containing at least 1000 bp of *H. humuli* major rDNA unit were assembled by Flye 2.8 (Kolmogorov et al., 2019) using minimal overlap 8 kbp. The annotation of MEs was done by RepeatMasker 4.1.2-p1 (Smit et al., 2013) protein-based masking. Tandem repeats were identified based on self Dotplot implemented in Geneious 11.1.5. Consensus sequences of all identified ME fragments together with major rDNA unit were mapped to individual rDNA bearing nanopore reads using minimap2 (Li, 2018) with appropriate pre-set. The presence and relative localization of individual elements was evaluated via R script (R version 4.0.3 in Rstudio version 1.4.1103). Only regions with mapping quality at least 20 were considered.

Phymatopus californicus gDNA was sequenced on Oxford Nanopore platform in Novogene Co.,Ltd. PacBio HiFi reads of *I. io* (project PRJEB42130) and *A. urticae* (project PRJEB42112) were obtain through the Darwin Tree of Life project (http://www.darwintreeoflife.org). PacBio CLR data were obtained from Sequence Read Archive (SRA) database (*S. frugiperda* SRR12642577; *L. dispar* SRR13505170-6, SRR13505182-3, and SRR13505187; *P.xylostella* SRR13530960). Further, the reads were processed same as in *H. humuli* except for the HiFi reads, which were not quality filtered.

Similar approach to detect rDNA and associated repetitive DNA was used also in *A. urticae* chromosomal level genome assembly (Bishop et al., 2021) (ENA acc. No. PRJEB41896).

Coverage analysis

Coverage analysis was done by aligning genomic Illumina sequencing reads from *H. humuli, I. io*, and *A. urticae* to consensus sequences, which were generated by overlapping the contigs from RE in Geneious 11.1.5 or by Flye 2.8 assembler, using Bowtie2 aligner (Langmead et al., 2019; Langmead and Salzberg, 2012). Coverage values were obtained using samtools depth (v 1.10) (Li et al., 2009) and plotted using a script in R (R version 4.1.0 in Rstudio Workbench Version 1.4.1717-3). Mean coverage of defined annotation blocks as seen in Figure 3 was computed using R and is in Suppl. Tables 3.

Data availability

Sequencing data generated in this study was deposited in the NCBI Sequence read archive under Bioproject reg. no. PRJNA737195. Long reads bearing rDNA of species under study, assemblies of their rDNA units and R codes used for analyses were deposited in the Dryad Digital Repository under doi: 10.5061/dryad.gmsbcc2qj.

Acknowledgement

We would like to thank our colleagues Anna Voleníková, František Marec, Leonela Carabajal-Paladino, Magda Zrzavá, Vladimír Beneš, and Marie Korchová for their help and extensive support. We would also like to show our gratitude to the Darwin Tree of Life project for allowing us to use its HiFi PacBio data. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures". Computational

resources were also provided by the ELIXIR-CZ project (LM2015047), part of the international ELIXIR infrastructure.

Funding

This research was supported by the grant No. 20-20650Y of the Czech Science Foundation to PN.

References

Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, Välimäki N, Paulin L, Kvist J, Wahlberg N, Tanskanen J, Hornett EA, Ferguson LC, Luo S, Cao Z, de Jong MA, Duplouy A, Smolander O-P, Vogel H, McCoy RC, Qian K, Chong WS, Zhang Q, Ahmad F, Haukka JK, Joshi A, Salojärvi J, Wheat CW, Grosse-Wilde E, Hughes D, Katainen R, Pitkänen E, Ylinen J, Waterhouse RM, Turunen M, Vähärautio A, Ojanen SP, Schulman AH, Taipale M, Lawson D, Ukkonen E, Mäkinen V, Goldsmith MR, Holm L, Auvinen P, Frilander MJ, Hanski I. 2014. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat Commun* **5**:4737. doi:10.1038/ncomms5737

Barbosa P, de Oliveira LA, Pucci MB, Santos MH, Moreira-Filho O, Vicari MR, Nogaroto V, de Almeida MC, Artoni RF. 2015. Identification and chromosome mapping of repetitive elements in the *Astyanax scabripinnis* (Teleostei: Characidae) species complex. *Genetica* **143**:55–62. doi:10.1007/s10709-014-9813-2

Baumgärtner L, Paiz LM, Zawadzki CH, Margarido VP, Portela Castro ALB. 2014. Heterochromatin polymorphism and physical mapping of 5S and 18S ribosomal DNA in four populations of *Hypostomus strigaticeps* (Regan, 1907) from the Paraná river basin, Brazil: Evolutionary and environmental correlation. *Zebrafish* **11**:479–487. doi:10.1089/zeb.2014.1028

Bedo DG. 1984. Karyotypic and chromosome banding studies of the potato tuber moth, *Phthorimaea operculella* (Zeller) (Lepidoptera, Gelechiidae). *Can J Genet Cytol* **26**:141–145. doi:10.1139/g84-024

Belser C, Baurens F-C, Noel B, Martin G, Cruaud C, Istace B, Yahiaoui N, Labadie K, Hřibová E, Doležel J, Lemainque A, Wincker P, D'Hont A, Aury J-M. 2021. Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. *Commun Biol* **4**:1047. doi:10.1038/s42003-021-02559-3

Bishop G, Ebdon S, Lohse K, Vila R, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium. 2021. The genome sequence of the small tortoiseshell butterfly, *Aglais urticae* (Linnaeus, 1758). *Wellcome Open Res* **6**:233. doi:10.12688/wellcomeopenres.17197.1

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170

Bueno D, Palacios-Gimenez OM, Cabral-de-Mello DC. 2013. Chromosomal mapping of repetitive DNAs in the grasshopper *Abracris flavolineata* reveal possible ancestry of the B chromosome and H3 histone spreading. *PLoS ONE* **8**:e66532. doi:10.1371/journal.pone.0066532

Cabral-de-Mello DC, Oliveira SG, de Moura RC, Martins C. 2011. Chromosomal organization of the 18S and 5S rRNAs and histone H3 genes in Scarabaeinae coleopterans: insights into the evolutionary dynamics of multigene families and heterochromatin. *BMC Genetics* **12**:88. doi:10.1186/1471-2156-12-88

Cabral-de-Mello DC, Zrzavá M, Kubíčková S, Rendón P, Marec F. 2021. The role of satellite DNAs in genome architecture and sex chromosome evolution in Crambidae moths. *Front Genet* **12**:661417. doi:10.3389/fgene.2021.661417

Carabajal Paladino LZ, Nguyen P, Šíchová J, Marec F. 2014. Mapping of single-copy genes by TSA-FISH in the codling moth, *Cydia pomonella*. *BMC Genet* **15**:S15. doi:10.1186/1471-2156-15-S2-S15

Carabajal Paladino LZ, Provazníková I, Berger M, Bass C, Aratchige NS, López SN, Marec F, Nguyen P. 2019. Sex chromosome turnover in moths of the diverse superfamily Gelechioidea. *Genome Biol Evol* **11**:1307–1319. doi:10.1093/gbe/evz075

Cheng T, Wu J, Wu Y, Chilukuri RV, Huang L, Yamamoto K, Feng L, Li W, Chen Z, Guo H, Liu J, Li S, Wang X, Peng L, Liu D, Guo Y, Fu B, Li Z, Liu C, Chen Y, Tomar A, Hilliou F, Montagné N, Jacquin-Joly E, d'Alençon E, Seth RK, Bhatnagar RK, Jouraku A, Shiotsuki T, Kadono-Okuda K, Promboon A, Smagghe G, Arunkumar KP, Kishino H, Goldsmith MR, Feng Q, Xia Q, Mita K. 2017. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nat Ecol Evol* **1**:1747–1756. doi:10.1038/s41559-017-0314-4

da Silva M, Barbosa P, Artoni RF, Feldberg E. 2016. Evolutionary dynamics of 5S rDNA and recurrent association of transposable elements in electric fish of the family Gymnotidae (Gymnotiformes): The case of *Gymnotus mamiraua*. *Cytogenet Genome Res* **149**:297–303. doi:10.1159/000449431

Dalíková Martina, Zrzavá M, Hladová I, Nguyen P, Šonský I, Flegrová M, Kubíčková S, Voleníková A, Kawahara AY, Peters RS, Marec F. 2017a. New insights into the evolution of the W chromosome in Lepidoptera. *J Hered* **108**:709–719. doi:10.1093/jhered/esx063

Dalíková M., Zrzavá M, Kubíčková S, Marec F. 2017b. W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). *Chromosome Res* **25**:241–252. doi:10.1007/s10577-017-9558-8

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **34**:2666–2669. doi:10.1093/bioinformatics/bty149

de Sene VF, Pansonato-Alves JC, Ferreira DC, Utsunomia R, Oliveira C, Foresti F. 2015. Mapping of the retrotransposable elements *Rex1* and *Rex3* in chromosomes of *Eigenmannia* (Teleostei, Gymnotiformes, Sternopygidae). *Cytogenet Genome Res* **146**:319–324. doi:10.1159/000441465

Eickbush TH, Eickbush DG. 2015. Integration, regulation, and long-term stability of R2 retrotransposons. *Microbiol Spectr* **3**:2. doi:10.1128/microbiolspec.MDNA3-0011-2014

Eickbush TH, Eickbush DG. 2007. Finely orchestrated movements: Evolution of the ribosomal RNA genes. *Genetics* **175**:477–485. doi:10.1534/genetics.107.071399

Elliott TA, Stage DE, Crease TJ, Eickbush TH. 2013. In and out of the rRNA genes: characterization of *Pokey* elements in the sequenced *Daphnia* genome. *Mobile DNA* **4**:20. doi:10.1186/1759-8753-4-20

Ferretti ABSM, Ruiz-Ruano FJ, Milani D, Loreto V, Martí DA, Ramos E, Martins C, Cabral-de-Mello DC. 2019. How dynamic could be the 45S rDNA cistron? An intriguing variability in a grasshopper species revealed by integration of chromosomal and genomic data. *Chromosoma* **128**:165–175. doi:10.1007/s00412-019-00706-8

Fiore-Donno AM, Kamono A, Meyer M, Schnittler M, Fukui M, Cavalier-Smith T. 2012. 18S rDNA phylogeny of *Lamproderma* and allied genera (Stemonitales, Myxomycetes, Amoebozoa). *PLoS ONE* **7**:e35359. doi:10.1371/journal.pone.0035359

Fujiwara H, Ogura T, Takada N, Miyajima N, Ishikawa H, Maekawa H. 1984. Introns and their flanking sequences of *Bombyx mori* rDNA. *Nucl Acids Res* **12**:6861–6869. doi:doi: 10.1093/nar/12.17.6861

Fuková I, Nguyen P, Marec F. 2005. Codling moth cytogenetics: Karyotype, chromosomal location of rDNA, and molecular differentiation of sex chromosomes. *Genome* **48**:1083–1092. doi:10.1139/g05-063

García-Souto D, Pérez-García C, Kendall J, Pasantes J. 2016. Molecular cytogenetics in trough shells (Mactridae, Bivalvia): Divergent GC-rich heterochromatin content. *Genes* **7**:47. doi:10.3390/genes7080047

Goldman ASH, Lichten M. 1996. The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. *Genetics* **144**:43–55. doi:10.1093/genetics/144.1.43

Grozdanov P, Georgiev O, Karagyozov L. 2003. Complete sequence of the 45-kb mouse ribosomal DNA repeat: analysis of the intergenic spacer. *Genomics* **82**:637–643. doi:10.1016/S0888-7543(03)00199-X

Havlová K, Dvořáčková M, Peiro R, Abia D, Mozgová I, Vansáčová L, Gutierrez C, Fajkus J. 2016. Variation of 45S rDNA intergenic spacers in *Arabidopsis thaliana*. *Plant Mol Biol* **92**:457–471. doi:10.1007/s11103-016-0524-1

Ingle J, Timmis JN, Sinclair J. 1975. The relationship between satellite deoxyribonucleic acid, ribosomal ribonucleic acid gene redundancy, and genome size in plants. *Plant Physiol* **55**:496–501. doi:10.1104/pp.55.3.496

Jakubczak JL, Burke WD, Eickbush TH. 1991. Retrotransposable elements RI and R2 interrupt the rRNA genes of most insects. *PNAS* **88**:3295-3299,. doi:10.1073/pnas.88.8.3295

Kato A, Albert P, Vega J, Kato A, Albert P, Vega J, Birchler J. 2006. Sensitive fluorescence *in situ* hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. *Biotechnic Histochem* **81**:71–78. doi:10.1080/10520290600643677

Kierzek E, Christensen SM, Eickbush TH, Kierzek R, Turner DH, Moss WN. 2009. Secondary structures for 5' regions of R2 retrotransposon RNAs reveal a novel conserved pseudoknot and regions that evolve under different constraints. *J Mol Biol* **390**:428–442. doi:10.1016/j.jmb.2009.04.048

Kobayashi T. 2011. Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell Mol Life Sci* **68**:1395–1403. doi:10.1007/s00018-010-0613-2

Kojima KK, Fujiwara H. 2005. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol Biol Evol* **22**:2157–2165. doi:10.1093/molbev/msi210

Kolmogorov M, Yuan J, Lin Y. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**:540–546. doi:10.1038/s41587-019-0072-8

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359. doi:10.1038/nmeth.1923

Langmead B, Wilks C, Antonescu V, Charles R. 2019. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**:421–432. doi:10.1093/bioinformatics/bty648

Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100. doi:10.1093/bioinformatics/bty191

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352

Lohe AR, Roberts PA. 2000. Evolution of DNA in heterochromatin: The *Drosophila melanogaster* sibling species subgroup as a resource. *Genetica* **109**:125–130. doi:10.1023/A:1026588217432

Long EO, Dawid IB. 1980. Repeated genes in eukaryotes. *Annu Rev Biochem* **49**:727–764. doi:49.070180.003455

Long EO, Dawid IB. 1979. Expression of ribosomal DNA insertions in *Drosophila melanogaster*. *Cell* **18**:1185–1196. doi:10.1016/0092-8674(79)90231-9

Lohse K, Mackintosh A, Vila R, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium .2021 The genome sequence of the European peacock butterfly, *Aglais io* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]. *Wellcome Open Res* **6**:258 doi:10.12688/wellcomeopenres.17204.1

Lu Y, Kochert GD, Isenhour DJ, Adang MJ. 1994. Molecular characterization of a strain-specific repeated DNA sequence in the fall armyworm *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Insect Molecular Biology* **3**:123–130. doi:10.1111/j.1365-2583.1994.tb00159.x

Lukhtanov V. 2015. The blue butterfly *Polyommatus (Plebicula) atlanticus* (Lepidoptera, Lycaenidae) holds the record of the highest number of chromosomes in the non-polyploid eukaryotic organisms. *Comparat Cytogenet* **9**:683–690. doi:10.3897/CompCytogen.v9i4.5760

Mahendran B, Acharya C, Dash R, Ghosh SK, Kundu SC. 2006. Repetitive DNA in tropical tasar silkworm *Antheraea mylitta*. *Gene* **370**:51–57. doi:10.1016/j.gene.2005.11.010

Mandrioli M, Manicardi GC, Marec F. 2003. Cytogenetic and molecular characterization of the MBSAT1 satellite DNA in holokinetic chromosomes of the cabbage moth, *Mamestra brassicae* (Lepidoptera). *Chromosome Res* **11**:51–56. doi:10.1023/a:1022058032217

Martí E, Milani D, Bardella VB, Albuquerque L, Song H, Palacios-Gimenez OM, Cabral-de-Mello DC. 2021. Cytogenomic analysis unveils mixed molecular evolution and recurrent chromosomal rearrangements shaping the multigene families on *Schistocerca* grasshopper genomes. *Evolution* **75**:2027-2041. doi:10.1111/evo.14287

McKinlay A, Fultz D, Wang F, Pikaard CS. 2021. Targeted enrichment of rRNA gene tandem arrays for ultra-long sequencing by selective restriction endonuclease digestion. *Front Plant Sci* **12**:656049. doi:10.3389/fpls.2021.656049

Mediouni J, Fuková I, Frydrychová R, Dhouibi MH, Marec F. 2004. Karyotype, sex chromatin and sex chromosome differentiation in the carob moth, *Ectomyelois ceratoniae* (Lepidoptera: Pyralidae). *Caryologia* **57**:184–194. doi:10.1080/00087114.2004.10589391

Muirhead CA, Presgraves DC. 2021. Satellite DNA-mediated diversification of a sex-ratio meiotic drive gene family in *Drosophila*. *Nat Ecol Evol* **5**:1604–1612. doi:10.1038/s41559-021-01543-8

Nelson JO, Slicko A, Yamashita YM. 2021. The retrotransposon R2 maintains *Drosophila* ribosomal DNA repeats. doi:10.1101/2021.07.12.451825

Nguyen P, Sahara K, Yoshido A, Marec F. 2010. Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). *Genetica* **138**:343–354. doi:10.1007/s10709-009-9424-5

Noller HF, Lancaster L, Mohan S, Zhou J. 2017. Ribosome structural dynamics in translocation: yet another functional role for ribosomal RNA. *Quart Rev Biophys* **50**:e12. doi:10.1017/S0033583517000117

Novák P, Neumann P, Macas J. 2010. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **11**:378. doi:10.1186/1471-2105-11-378

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29**:792–793. doi:10.1093/bioinformatics/btt054

Palacios-Gimenez OM, Castillo ER, Martí DA, Cabral-de-Mello DC. 2013. Tracking the evolution of sex chromosome systems in Melanoplinae grasshoppers through chromosomal mapping of repetitive DNA sequences. *BMC Evol Biol* **13**:167. doi:10.1186/1471-2148-13-167

Perumal S, Waminal NE, Lee Jonghoon, Lee Junki, Choi B-S, Kim HH, Grandbastien M-A, Yang T-J. 2017. Elucidating the major hidden genomic components of the A, C, and AC genomes and their influence on *Brassica* evolution. *Sci Rep* **7**:17986. doi:10.1038/s41598-017-18048-9

Poletto AB, Ferreira IA, Cabral-de-Mello DC, Nakajima RT, Mazzuchelli J, Ribeiro HB, Venere PC, Nirchio M, Kocher TD, Martins C. 2010. Chromosome differentiation patterns during cichlid fish evolution. *BMC Genet* **11**:50. doi:10.1186/1471-2156-11-50

Prins JD, Saitoh K. 2003. Karyology and sex determination In: Kükenthal W, editor. Band 4: Arthropoda, 2 Hälfte: Insecta, Lepidoptera, Moths and Butterflies, Teilband/Part 36, Vol 2: Morphology, Physiology, and Development. Berlin, Boston: DE GRUYTER. doi:10.1515/9783110893724.449

Prokopowich CD, Gregory TR, Crease TJ. 2003. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46**:48–50. doi:10.1139/g02-103

Proux-Wéra E, Byrne KP, Wolfe KH. 2013. Evolutionary mobility of the ribosomal DNA array in yeasts. *Genome Biol Evol* **5**:525–531. doi:10.1093/gbe/evt022

Provazníková I, Hejníčková M, Visser S, Dalíková M, Carabajal Paladino LZ, Zrzavá M, Voleníková A, Marec F, Nguyen P. 2021. Large-scale comparative analysis of cytogenetic markers across Lepidoptera. *Sci Rep* **11**:12214. doi:10.1038/s41598-021-91665-7

Raskina O, Barber JC, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: Speciation-related events in plant genomes. *Cytogenet Genome Res* **120**:351–357. doi:10.1159/000121084

Raskina O, Belyayev A, Nevo E. 2004. Activity of the *En/Spm*-like transposons in meiosis as a base for chromosome repatterning in a small, isolated, peripheral population of *Aegilops speltoides* Tausch. *Chromosome Res* **12**:153–161. doi:10.1023/B:CHRO.0000013168.61359.43

Robinson R. 1971. Lepidoptera Genetics. Oxford: Pergamon Press.

Roiha H, Miller J, Woods L. 1981. Arrangements and rearrangements of sequences flanking the two types of rDNA insertion in *D. melanogaster*. *Nature* **290**:749–754. doi:10.1038/290749a0

Ruiz-Herrera A, Nergadze SG, Santagostino M, Giulotto E. 2008. Telomeric repeats far from the ends: Mechanisms of origin and role in evolution. *Cytogenet Genome Res* **122**:219–228. doi:10.1159/000167807

Scacchetti PC, Alves de Oliveira JC, Utsunomia R, Claro FL, de Almeida Toledo LF, Oliveira C, Foresti F. 2012. Molecular characterization and physical mapping of two classes of 5S rDNA in the genomes of *Gymnotus sylvius* and *G. inaequilabiatus* (Gymnotiformes, Gymnotidae). *Cytogenet Genome Res* **136**:131–137. doi:10.1159/000335658

Scali V, Coluccia E, Deidda F, Lobina C, Deiana AM, Salvadori S. 2016. Co-localization of ribosomal and telomeric sequences in *Leptynia* (Insecta: Phasmatodea). *Ital J Zool* **83**:285–290. doi:10.1080/11250003.2016.1219403

Sember A, Bohlen J, Šlechtová V, Altmanová M, Pelikánová Š, Ráb P. 2018. Dynamics of tandemly repeated DNA sequences during evolution of diploid and tetraploid botiid loaches (Teleostei: Cobitoidea: Botiidae). *PLoS ONE* **13**:e0195054. doi:10.1371/journal.pone.0195054

Šíchová J, Nguyen P, Dalíková M, Marec F. 2013. Chromosomal evolution in tortricid moths: Conserved karyotypes with diverged features. *PLoS ONE* **8**:e64520. doi:10.1371/journal.pone.0064520

Šíchová J, Ohno M, Dincă V, Watanabe M, Sahara K, Marec F. 2016. Fissions, fusions, and translocations shaped the karyotype and multiple sex chromosome constitution of the northeast-Asian wood white butterfly, *Leptidea amurensis*. *Biol J Linn Soc* **118**:457–471. doi:10.1111/bij.12756

Šíchová J, Voleníková A, Dincă V, Nguyen P, Vila R, Sahara K, Marec F. 2015. Dynamic karyotype evolution and unique sex determination systems in Leptidea wood white butterflies. *BMC Evol Biol* **15**:89. doi:10.1186/s12862-015-0375-4

Silva DMZ de A, Pansonato-Alves JC, Utsunomia R, Araya-Jaime C, Ruiz-Ruano FJ, Daniel SN, Hashimoto DT, Oliveira C, Camacho JPM, Porto-Foresti F, Foresti F. 2014. Delimiting the origin of a B chromosome by FISH mapping, chromosome painting and DNA sequence analysis in *Astyanax paranae* (Teleostei, Characiformes). *PLoS ONE* **9**:e94896. doi:10.1371/journal.pone.0094896

Sims J, Sestini G, Elgert C, von Haeseler A, Schlögelhofer P. 2021. Sequencing of the *Arabidopsis* NOR2 reveals its distinct organization and tissue-specific rRNA ribosomal variants. *Nat Commun* **12**:387. doi:10.1038/s41467-020-20728-6

Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0.

Sochorová J, Garcia S, Gálvez F, Symonová R, Kovařík A. 2018. Evolutionary trends in animal ribosomal DNA loci: introduction to a new online database. *Chromosoma* **127**:141–150. doi:10.1007/s00412-017-0651-8

Sproul JS, Barton LM, Maddison DR. 2020. Repetitive DNA profiles reveal evidence of rapid genome evolution and reflect species boundaries in ground beetles. *Systematic Biology* **69**:1137–1148. doi:10.1093/sysbio/syaa030

Symonová R. 2019. Integrative rDNAomics—Importance of the oldest repetitive fraction of the eukaryote genome. *Genes* **10**:345. doi:10.3390/genes10050345

Symonová R, Majtánová Z, Sember A, Staaks GB, Bohlen J, Freyhof J, Rábová M, Ráb P. 2013. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol Biol* **13**:42. doi:10.1186/1471-2148-13-42

Van Nieukerken EJ, Kaila L, Kitching IJ, Kristensen NP, Lees DC, Minet J, Mitter C, Mutanen M, Regier JC, Simonsen TJ, Wahlberg N, Yen S-H, Zahiri R, Adamski D, Baixeras J, Bartsch D, Bengtsson BÅ, Brown JW, Bucheli SR, Davis DR, Prins JD, Prins WD, Epstein ME, Gentili-Poole P, Gielis C, Hättenschwiler P, Hausmann A, Holloway JD, Kallies A, Karsholt O, Kawahara AY, Koster SJC, Kozlov MV, Lafontaine JD, Lamas G, Landry J-F, Lee S, Nuss M, Park K-T, Penz C, Rota J, Schintlmeister A, Schmidt BC, Sohn J-C, Solis MA, Tarmann GM, Warren AD, Weller S, Yakovlev RV, Zolotuhin VV, Zwick A. 2011. Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa* **3148**:212. doi:10.11646/zootaxa.3148.1.41

Van't Hof AE, Nguyen P, Dalíková M, Edmonds N, Marec F, Saccheri IJ. 2013. Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a model of industrial melanism. *Heredity* **110**:283–295. doi:10.1038/hdy.2012.84

Věchtová P, Dalíková M, Sýkorová M, Žurovcová M, Füssy Z, Zrzavá M. 2016. CpSAT-1, a transcribed satellite sequence from the codling moth, *Cydia pomonella*. *Genetica* **144**:385–395. doi:10.1007/s10709-016-9907-0

Vershinina AO, Anokhin BA, Lukhtanov VA. 2015. Ribosomal DNA clusters and telomeric (TTAGG)n repeats in blue butterflies (Lepidoptera, Lycaenidae) with low and high chromosome numbers. *CCG* **9**:161–171. doi:10.3897/CompCytogen.v9i2.4715

Vítková M, Fuková I, Kubíčková S, Marec F. 2007. Molecular divergence of the W chromosomes in pyralid moths (Lepidoptera). *Chromosome Res* **15**:917–930. doi:10.1007/s10577-007-1173-7

Vondrak T, Oliveira L, Novák P, Koblížková A, Neumann P, Macas J. 2021. Complex sequence organization of heterochromatin in the holocentric plant *Cuscuta europaea* elucidated by the computational analysis of nanopore reads. *Comput Struct Biotechnol* **19**:2179–2189. doi:10.1016/j.csbj.2021.04.011

Wolf KW, Novak K, Marec F. 1997. Kinetic organization of metaphase I bivalents in spermatogenesis of Lepidoptera and Trichoptera species with small chromosome numbers. *Heredity* **79**:35—143.

Wu S, Xiong J, Yu Y. 2015. Taxonomic resolutions based on 18S rRNA genes: A case study of subclass Copepoda. *PLoS ONE* **10**:e0131498. doi:10.1371/journal.pone.0131498

Xiong Y, Burke WD, Jakubczak JL, Eickbush TH. 1988. Ribosomal DNA insertion elements R1Bm and R2Bm can transpose in a sequence specific manner to locations outside the 28S genes. *Nucleic Acids Res* **16**:10561–10573. doi:10.1093/nar/16.22.10561

Yano CF, Merlo MA, Portela-Bens S, Cioffi M de B, Bertollo LAC, Santos-Júnior CD, Rebordinos L. 2020. Evolutionary dynamics of multigene families in *Triportheus* (Characiformes, Triportheidae): A transposon mediated mechanism? *Front Mar Sci* **7**:6. doi:10.3389/fmars.2020.00006

Yasukochi Y, Tanaka-Okuyama M, Kamimura M, Nakano R, Naito Y, Ishikawa Y, Sahara K. 2011. Isolation of BAC clones containing conserved genes from libraries of three distantly related moths: A useful resource for comparative genomics of Lepidoptera. *J Biomed Biotechnol* **2011**:1–6. doi:10.1155/2011/165894

Zhou J, Eickbush MT, Eickbush TH. 2013. A Population genetic model for the maintenance of R2 retrotransposons in rRNA gene loci. *PLoS Genet* **9**:e1003179. doi:10.1371/journal.pgen.1003179

Zrzavá M, Hladová I, Dalíková M, Šíchová J, Õunap E, Kubíčková S, Marec F. 2018. Sex chromosomes of the iconic moth *Abraxas grossulariata* (Lepidoptera, Geometridae) and its congener *A. sylvata*. *Genes* **9**:279. doi:10.3390/genes9060279

Figures

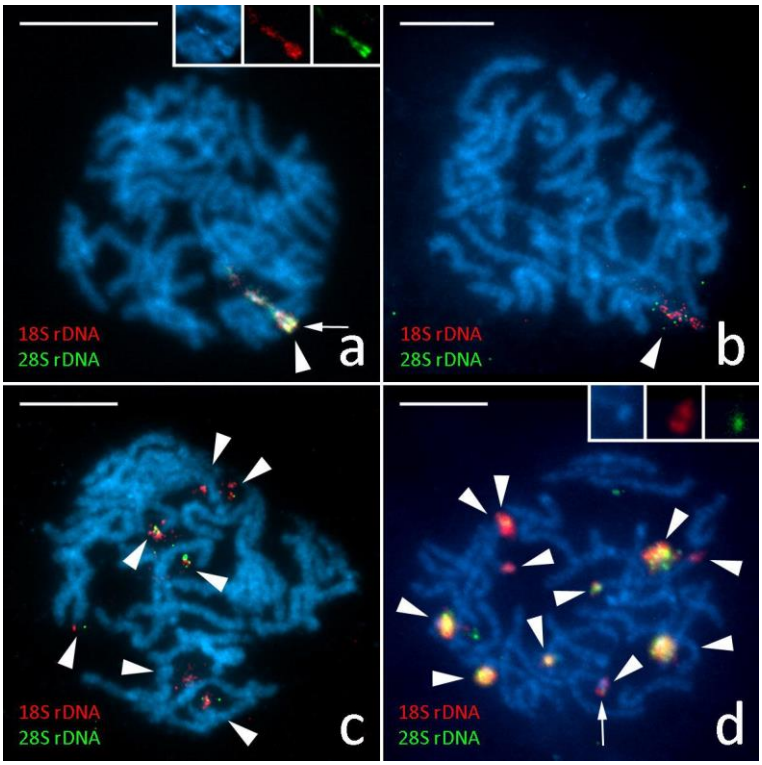

Figure 1: Major rDNA clusters detected by fluorescence in situ hybridization (FISH) on pachytene nuclei of studied species. The 18S rDNA probe (red) and 28S rDNA probe (green), chromosomes are counterstained with DAPI (blue). Major rDNA clusters indicated by arrowheads. **a** – male pachytene nucleus of *H. humuli* (Hepialidae) with detail of the major rDNA cluster colocalizing with heterochromatin blocks in the inset, **b** – male pachytene nucleus of *C. ohridella* (Gracillariidae), **c** – male pachytene nucleus of *A. urticae*(Nymphalidae), **d** – female pachytene nucleus of *I. io* (Nymphalidae) with detail of one of major rDNA clusters colocalizing with a small heterochromatin block in the inset. * clusters colocalizing with DAPI positive heterochromatin. Scale 10 µm

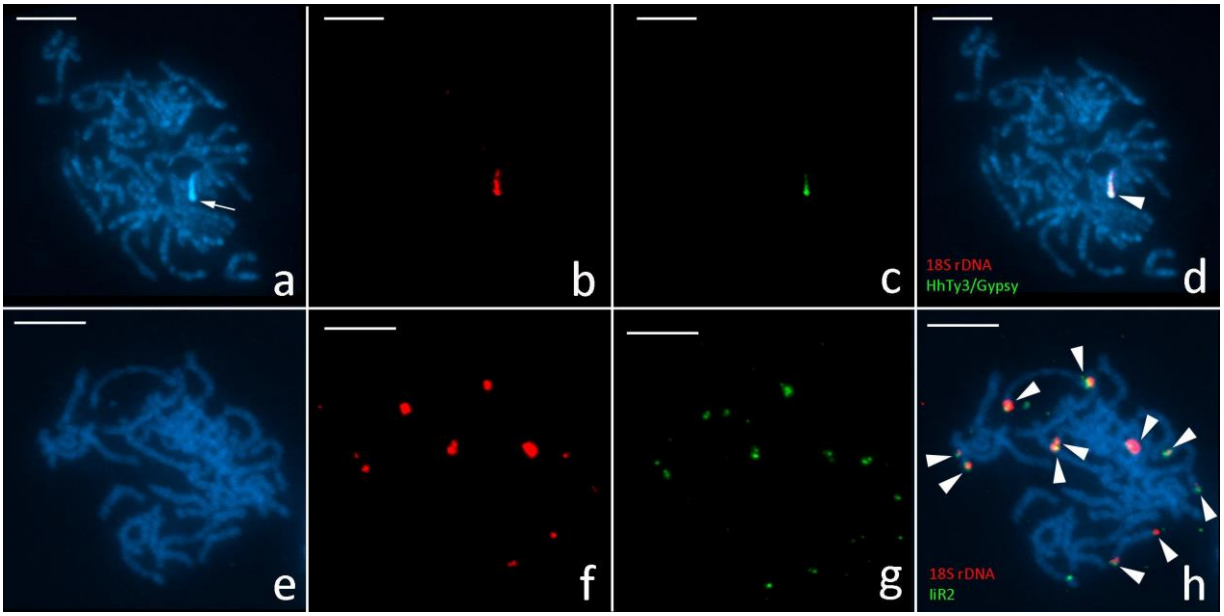

Figure 2: Co-localization of major rDNA and ME sequences of interest as detected by fluorescence *in situ* hybridization (FISH) on pachytene nuclei of studied species. 18S rDNA probe in red (**b, d, f, h**) and HhTy3/GypsyA (**c, d**) and IiR2 probes (**g, h**) in green, chromosomesare counterstained with DAPI (blue; **a**, **d, e, h**). Hybridization signals indicated by arrowhead.**a-d** – male pachytene nucleus of *H. humuli* (Hepialidae), **e-h** – female pachytene nucleus of *I.io* (Nymphalidae). Arrow indicates DAPI positive heterochromatin. Scale 10 µm

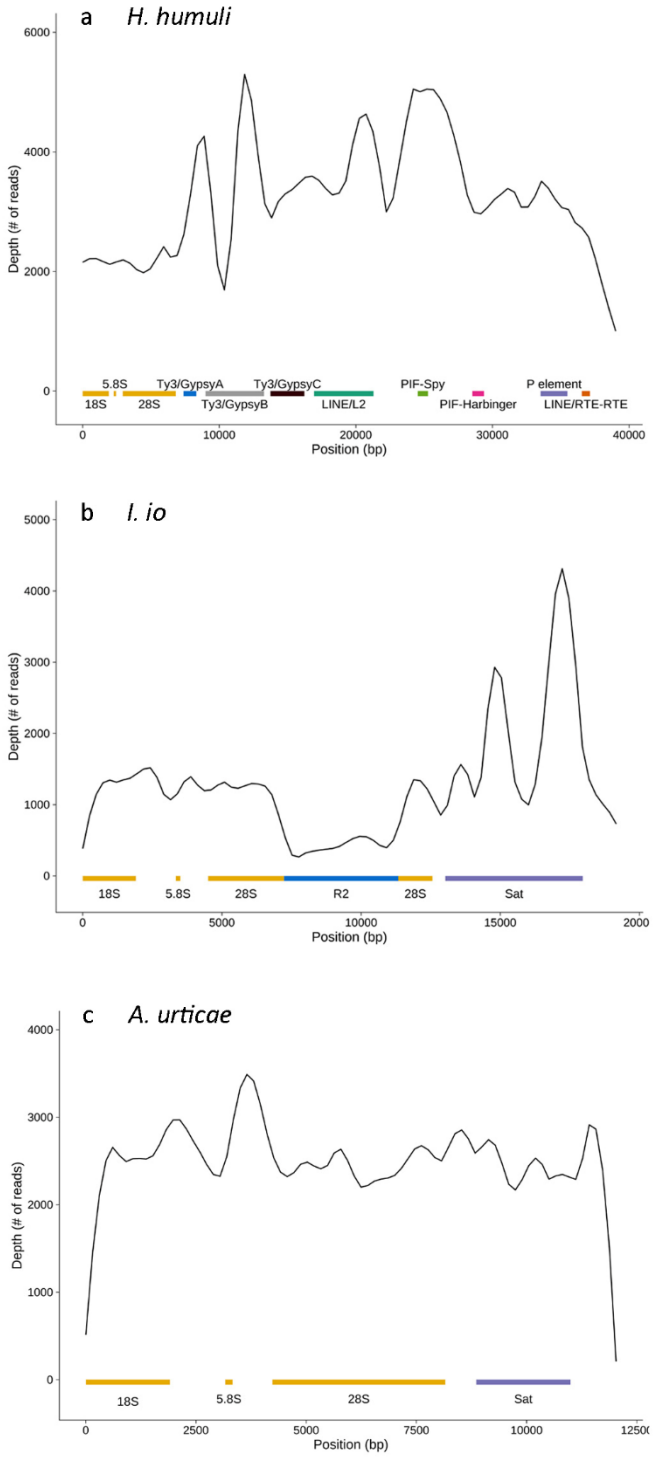

Figure 3: Coverage plot of rDNA units in *H. humuli* (**a**), *I. io* (**b**), and *A. urticae* (**c**). Smoothed (LOESS) counts of aligned sequencing reads for each nucleotide position of the major rDNA cluster. Coloured bars on the bottom represent regions of repetitive elements and position of rDNA genes.

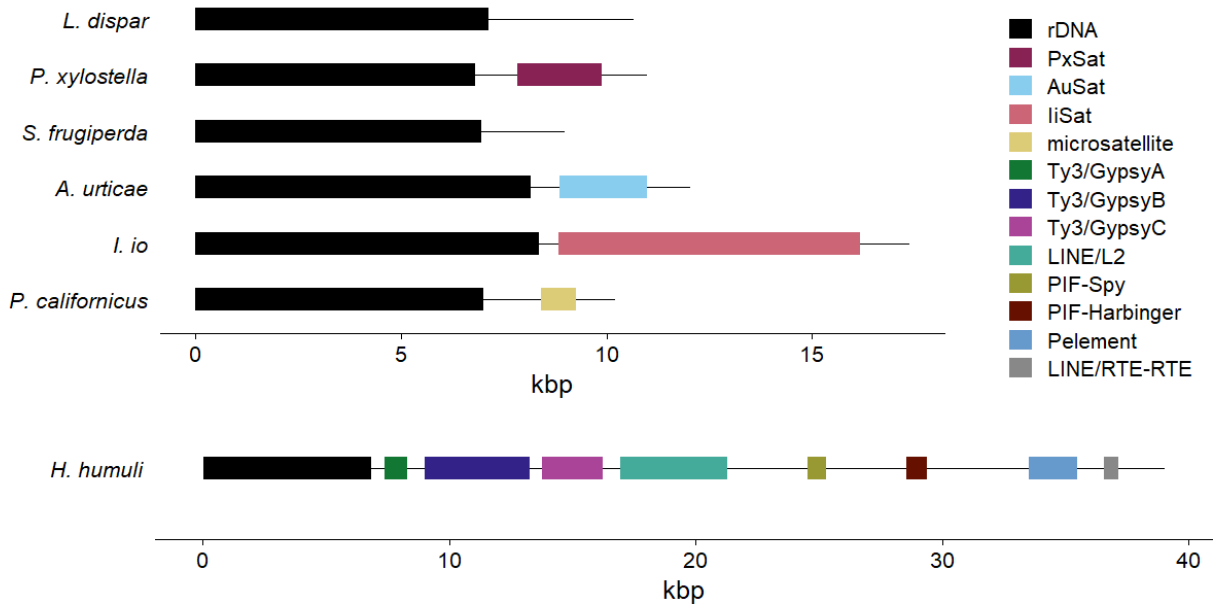

Figure 4: Schematic representation of most observed rDNA unit based on long read analysis. Black line represents the length of rDNA unit and coloured blocks position of rDNA genes and the repetitive elements.

Supplementary material

https://www.biorxiv.org/content/10.1101/2022.03.26.485928v1.supplementary-material

3.3. Chapter III

Sex chromosome turnover in moths of the diverse superfamily Gelechioidea.

Carabajal Paladino L.Z., Provazníková I., Berger M., Bass C., Aratchige N. S., López S.N., Marec F., Petr Nguyen P.

Genome Biology and Evolution (2019) 11: 1307–1319, doi: 10.1093/gbe/evz075

Abstract

Sex chromosomes play a central role in genetics of speciation and their turnover was suggested to promote divergence. In vertebrates, sex chromosome–autosome fusions resulting in neo-sex chromosomes occur frequently in male heterogametic taxa (XX/XY), but are rare in groups with female heterogamety (ZW/ZZ). We examined sex chromosomes of seven pests of the diverse lepidopteran superfamily Gelechioidea and confirmed the presence of neo-sex chromosomes in their karyotypes. Two synteny blocks, which correspond to autosomes 7 (LG7) and 27 (LG27) in the ancestral lepidopteran karyotype exemplified by the linkage map of Biston betularia (Geometridae), were identified as sex-linked in the tomato leafminer, Tuta absoluta (Gelechiidae). Testing for sex-linkage performed in other species revealed that while LG7 fused to sex chromosomes in a common ancestor of all Gelechioidea, the second fusion between the resulting neo-sex chromosome and the other autosome is confined to the tribe Gnoreschemini (Gelechiinae). Our data accentuate an emerging pattern of high incidence of neo-sex chromosomes in Lepidoptera, the largest clade with ZW/ZZ sex chromosome system, which suggest that the paucity of neo-sex chromosomes is not an intrinsic feature of female heterogamety. Furthermore, LG7 contains one of the major clusters of UDP-glucosyltransferases, which are involved in the detoxification of plant secondary metabolites. Sex chromosome evolution in Gelechioidea thus supports an earlier hypothesis postulating that lepidopteran sex chromosome-autosome fusions can be driven by selection for association of Z-linked preference or host-independent isolation genes with larval performance and thus can contribute to ecological specialization and speciation of moths.

Sex Chromosome Turnover in Moths of the Diverse Superfamily Gelechioidea

Leonela Z. Carabajal Paladino[1,2,†], Irena Provazníková[1,3,†], Madeleine Berger[4], Chris Bass[5], Nayanie S. Aratchige[6], Silvia N. López[7], František Marec[1], and Petr Nguyen[1,3,*]

[1]Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic

[2]The Pirbright Institute, Surrey, United Kingdom

[3]University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic

[4]Rothamsted Research, Department of Biointeractions and Crop Protection, Herts, United Kingdom

[5]University of Exeter, College of Life and Environmental Sciences, Biosciences, Penryn, Cornwall, United Kingdom

[6]Coconut Research Institute of Sri Lanka, Crop Protection Division, Bandirippuwa Estate, Lunuwila, Sri Lanka

[7]Instituto Nacional de Tecnología Agropecuaria, Instituto de Microbiología y Zoología Agrícola, Hurlingham, Buenos Aires, Argentina

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: petr.nguyen@prf.jcu.cz.

Accepted: April 5, 2019

Data deposition: This project has been deposited at GenBank under accession numbers MG265651, MG265652, MG265654–MG265661, MG265664–MG265670, MG265672–MG265675, MG265678–MG265690, and MG265692–MG265694.

Abstract

Sex chromosomes play a central role in genetics of speciation and their turnover was suggested to promote divergence. In vertebrates, sex chromosome–autosome fusions resulting in neo-sex chromosomes occur frequently in male heterogametic taxa (XX/XY), but are rare in groups with female heterogamety (WZ/ZZ). We examined sex chromosomes of seven pests of the diverse lepidopteran superfamily Gelechioidea and confirmed the presence of neo-sex chromosomes in their karyotypes. Two synteny blocks, which correspond to autosomes 7 (LG7) and 27 (LG27) in the ancestral lepidopteran karyotype exemplified by the linkage map of *Biston betularia* (Geometridae), were identified as sex-linked in the tomato leafminer, *Tuta absoluta* (Gelechiidae). Testing for sex-linkage performed in other species revealed that while LG7 fused to sex chromosomes in a common ancestor of all Gelechioidea, the second fusion between the resulting neo-sex chromosome and the other autosome is confined to the tribe Gnoreschemini (Gelechiinae). Our data accentuate an emerging pattern of high incidence of neo-sex chromosomes in Lepidoptera, the largest clade with WZ/ZZ sex chromosome system, which suggest that the paucity of neo-sex chromosomes is not an intrinsic feature of female heterogamety. Furthermore, LG7 contains one of the major clusters of UDP-glucosyltransferases, which are involved in the detoxification of plant secondary metabolites. Sex chromosome evolution in Gelechioidea thus supports an earlier hypothesis postulating that lepidopteran sex chromosome–autosome fusions can be driven by selection for association of Z-linked preference or host-independent isolation genes with larval performance and thus can contribute to ecological specialization and speciation of moths.

Key words: *Coleophora*, *Depressaria*, *Hofmannophila*, *Opisina*, *Phthorimaea*, *Sitotroga*.

Introduction

Sex chromosomes represent intriguing portions of the genome which play an important role in many evolutionary processes including sexual and intragenomic conflict and speciation (Masly and Presgraves 2007; Mank et al. 2014). Indeed, the formation of postzygotic isolation can be characterized by two empirical rules, both involving sex chromosomes, inferred from analyses of hybrid fitness. The first of these known as the large-X effect refers to the disproportionately large effect of the X chromosome compared with

autosomes in introgression analyses of hybrid incompatibilities (Masly and Presgraves 2007; Dufresnes et al. 2016). The second, Haldane's rule, which has proved to be one of the most robust generalizations in evolutionary biology, states that when in the F_1 offspring of two different animal races one sex is absent, rare, or sterile, that sex is the heterogametic sex (Haldane 1922; Delph and Demuth 2016).

It was shown that larger and more heteromorphic sex chromosomes were associated with faster evolution of postzygotic isolation (Turelli and Begun 1997; Lima 2014). Sex chromosome size can increase via sex chromosome–autosome fusions, which result in so-called neo-sex chromosomes. These have been suggested to promote divergence in fish (Kitano et al. 2009; Kitano and Peichel 2012), mammals (Graves 2016), and moths (Nguyen et al. 2013; Nguyen and Carabajal Paladino 2016), although little is known about their functional role in this process. Neo-sex chromosomes also provide insight into the evolution of animal sex chromosomes (Pala et al. 2012; Bachtrog 2013; Natri et al. 2013), which are much older than the sex chromosome systems examined in plants (Charlesworth 2015). To identify the evolutionary forces driving sex chromosome–autosome fusions, the occurrence of derived multiple sex chromosome systems was recently analyzed in vertebrates (Pokorná et al. 2014; Pennell et al. 2015). These analyses yielded a striking pattern of a higher incidence of fusions in male heterogametic (♀XX, ♂XY) than female heterogametic (♀WZ, ♂ZZ) taxa. Moreover, it was shown that Y–autosome fusions occur most frequently. Theoretical models suggested that a combination of two or more evolutionary forces, such as underdominance of the fusions, male-biased mutation rates for fusions, and female-biased reproductive sex ratio, is needed to explain the asymmetry between the Y and W chromosomes (Pennell et al. 2015; Kirkpatrick 2017).

Moths and butterflies (Lepidoptera), together with their sister order caddisflies (Trichoptera), constitute the most speciose lineage with female heterogamety. In their overview of 40 lepidopteran species with identified sex chromosomes, Traut et al. (2007) listed 12 moths with multiple sex chromosomes. Since then, more neo-sex chromosome systems have been reported in this order (Nguyen et al. 2013; Šíchová et al. 2013, 2015, 2016; Smith et al. 2016; Fraïsse et al. 2017; Mongue et al. 2017; Traut et al. 2017; Picq et al. 2018).

Some of the derived sex chromosome systems correspond to a conspicuously large sex chromosome pair (Nguyen et al. 2013; Šíchová et al. 2013; Mongue et al. 2017; Picq et al. 2018), which suggests that both W and Z sex chromosomes fused with an autosome. Similar large chromosome pairs were also observed in representatives of the families Pyralidae, Oecophoridae, and Gelechiidae with reduced chromosome numbers, but were considered autosomal fusion products (Ennis 1976). Carabajal Paladino et al. (2016), however, showed that the large chromosome pair corresponds to sex chromosomes in an invasive gelechiid pest, the tomato leafminer *Tuta absoluta* (Gelechiidae).

To test for the presence of neo-sex chromosomes in their genomes, we examined the karyotypes of several pests of the diverse superfamily Gelechioidea, which contains ~18,500 species (van Nieukerken et al. 2011) and comprises among others the above-mentioned Oecophoridae and Gelechiidae families. Our results confirmed a sex chromosome–autosome fusion, which occurred in a common ancestor of all three main lineages of Gelechioidea, namely the Gelechiid, Scythridid, and Depressariid assemblages (Sohn et al. 2016). A synteny block involved in the fusion was identified as an autosome homoeologous to the chromosome 7 of the ancestral karyotype represented by the peppered moth *Biston betularia* (Geometridae) (cf. Van't Hof 2013). Furthermore, we discovered another fusion between the neo-sex chromosomes and homoeologue of the *B. betularia* chromosome 27 within the tribe Gnorimoschemini (Gelechiinae). A potential role of the sex chromosome turnover in the divergence of Gelechioidea is discussed.

Materials and Methods

Insects

Representatives of five families within Gelechioidea were either obtained from laboratory stocks or collected from natural populations. A laboratory stock of the potato tuber moth, *Phthorimaea operculella* (Gelechiidae), was provided by the Atomic Energy Commission of Syria (Damascus, Syria). Larvae were reared on wax-coated potato slices as described in Saour and Makee (1997). Cultures of the Angoumois grain moth, *Sitotroga cerealella* (Gelechiidae), from the Instituto de Microbiología y Zoología Agrícola (IMYZA), Instituto Nacional de Tecnología Agropecuaria (INTA) (Buenos Aires, Argentina), and the Institute for Biological Control JKI, Federal Research Centre for Cultivated Plants (Darmstadt, Germany) were kept on wheat grains (Méndez et al. 2016). A laboratory colony of the tomato leafminer, *T. absoluta* (Gelechiidae), from IMYZA, INTA was maintained on potted tomato plants under the conditions detailed in Cagnotti et al. (2012). Specimens of the coconut black-headed caterpillar, *Opisina arenosella* (Xylorictidae), were obtained from the colony maintained on coconut leaflets at the Crop Protection Division of the Coconut Research Institute of Sri Lanka (Lunuwila, Sri Lanka). The larch case-bearer *Coleophora laricella* (Coleophoridae) and the brown house-moth *Hofmannophila pseudospretella* (Oecophoridae) were collected as larvae from wild populations in Levín (Lišov, Czech Republic). The dingy flat-body moth *Depressaria daucella* (Depressaridae) was collected as larvae and pupae in Slapy u Tábora (Tábor, Czech Republic). The material obtained in the field was immediately processed for its future analysis, and barcoded using a fragment of the *cytochrome c oxidase*

subunit I (*COI*) gene as described in Hebert et al. (2004). The sequences obtained were checked in the BOLD animal identification database (Ratnasingham and Hebert 2007) to confirm the identity of the specimens (for accession numbers of the sequences, see supplementary table S1, Supplementary Material online).

Processing of the Insects

Spread chromosome preparations were made from wing imaginal discs, testes, or ovaries of the last instar larvae of all species using the method of Traut (1976) with slight modifications detailed in Šíchová et al. (2013). For *D. daucella*, preparations were also made from ovaries of female pupae. The preparations were dehydrated in an ethanol series (70%, 80%, and 100%, 30 s each) and stored at −20 °C.

Nucleic acids were isolated from larvae or pupae. Given the size of the specimens, total RNA was recovered using the NucleoSpin RNA II (Macherey-Nagel, Düren, Germany) kit, RNA blue (Top-Bio, Prague, Czech Republic), or RNAzol (Sigma–Aldrich, St. Louis, MO). The first-strand cDNA was then synthesized by random or oligo-dT primed SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad, CA). Genomic DNA (gDNA) was extracted either by the NucleoSpin Tissue kit (Macherey-Nagel) or the MagAttract HMW DNA Kit (Qiagen, Hilden, Germany) and if needed, amplified by illustra GenomiPhi HY DNA Amplification Kit (GE Healthcare, Milwaukee, WI).

Fluorescence In Situ Hybridization Experiments

To identify sex chromosomes genomic in situ hybridization (GISH) was performed as described in Yoshido et al. (2005). Amplified male gDNA was fragmented by heating to 99 °C for 10 min in a TProfessional TRIO thermocycler (Biometra, Göttingen, Germany), and used as a species-specific competitor DNA (Šíchová et al. 2013). Female gDNA was labeled with fluorescein-12-dUTP (Jena Bioscience, Jena, Germany) using the nick translation protocol of Kato et al. (2006) with 3.5-h incubation at 15 °C. To accurately determine chromosome numbers, fluorescence in situ hybridization (FISH) with $(TTAGG)_n$ telomeric probes (tel-FISH) was performed either alone or in combination with GISH as described in Yoshido et al. (2005) and Šíchová et al. (2015). Unlabeled $(TTAGG)_n$ telomeric probes were prepared by nontemplate PCR according to Sahara et al. (1999) and labeled with Cy3-dUTP (Jena Bioscience) using the same nick translation protocol as above, but with 1-h incubation at 15 °C. For each slide, the hybridization mixture contained unlabeled fragmented male gDNA (3 μg) and female fluorescein-labeled gDNA (500 ng), and/or Cy3-labeled telomeric probe (200 ng), and sonicated salmon sperm DNA (25 μg). The preparations were counterstained with 0.5 mg/ml DAPI (4′,6-diamidino-2-phenylindole; Sigma–Aldrich) in antifade based on DABCO (1,4-diazabicyclo[2.2.2]octane; Sigma–Aldrich) (for composition, see Traut et al. 1999).

Preparations from FISH experiments were observed in a Zeiss Axioplan 2 microscope (Carl Zeiss, Jena, Germany) equipped with appropriate fluorescence filter sets. Black-and-white images were captured with an Olympus CCD monochrome camera XM10 equipped with cellSens 1.9 digital imaging software (Olympus Europa Holding, Hamburg, Germany). The images were pseudocolored and superimposed with Adobe Photoshop CS6 (Adobe Systems, San Jose, CA).

Screening for *T. absoluta* Sex-Linked Genes

The sex-linkage of selected genes was tested by means of quantitative PCR (qPCR) using male and female gDNA as template and autosomal gene as a reference (Nguyen et al. 2013; Dalíková et al. 2017). The selected genes were orthologous to markers for all the chromosomes of the ancestral karyotype represented by the *B. betularia* (Geometridae) linkage map (Van't Hof 2013) and the *Melitaea cinxia* (Nymphalidae) genome (Ahola et al. 2014) (supplementary table S2, Supplementary Material online). Primers were designed using available *T. absoluta* transcriptome sequences (Berger et al. 2016). The 1:1 (female:male) ratio of the used autosomal reference genes, *elongation factor 1 alpha* (*EF-1a*) and *acetylcholinesterase 1* (*Ace-1*) (using *Ace-1* as target and *EF-1a* as reference), and the 1:2 ratio of the Z-linked control gene *kettin* (*ket*) (using *Ace-1* as reference) were verified before analyzing other markers. The genes were tested in triplicates of three independent samples of both male and female gDNAs. Amplification efficiencies (*E*) of primer pairs were determined from the slope of the standard curve generated by plotting the threshold cycle (Ct) values against the log-concentrations of serial dilutions of male and female gDNAs. The female-to-male (F:M) ratio for each gene was calculated for each female as $[(1+E_{target})\hat{}(\text{Average\_Ct}_{target\_male} - \text{Ct}_{target\_female})] / [(1+E_{reference})\hat{}(\text{Average\_Ct}_{reference\_male} - \text{Ct}_{reference\_female})]$, and then compared with the expected values of 1 and 0.5 corresponding to autosomal position and sex-linkage, respectively, by means of one-sample *t*-test using R (R Core Team 2013) (supplementary table S3, Supplementary Material online). Composition of the reaction, cycling conditions, and sequences of forward and reverse primers are detailed in supplementary tables S3 and S4, Supplementary Material online.

Once these control genes were validated, the marker genes (supplementary table S2, Supplementary Material online) were analyzed using one biological replicate per sex with three technical replicates per gDNA sample. In this case, the F:M ratio was calculated using the delta delta Ct method as $2\hat{}[(\text{Ct}_{target\_female} - \text{Ct}_{reference\_female}) - (\text{Ct}_{target\_male} - \text{Ct}_{reference\_male})]$, which is a simplified version of the aforementioned formula that assumes $E = 1$ for all genes. The obtained

values were considered for analysis only if *ket* and/or the reference genes (*EF-1a* and *Ace-1*) compared with each other provided the expected and previously corroborated 0.5 and 1 values, respectively. The experiments were carried out at least three times, using both reference genes, *EF-1a* and *Ace-1*. All reactions were performed in a final volume of 25 µl using SYBR Premix Ex *Taq* II (Perfect Real Time) (TaKaRa, Otsu, Japan) and a final concentration of primers of 0.2 mM for both the target and reference genes (except for *pixie ATP-binding cassette subfamily E member 1* [*Pix*] when a final concentration of 0.3 mM was used for the primers of the target gene). The cycling conditions included an initial denaturation at 95 °C for 3 min, then 45 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s, a final denaturation of 95 °C for 15 s, and then an increase of temperature from 65 to 95 °C with increments of 0.5 °C for 5 s for the generation of melting curves. The sequences of forward and reverse primers are detailed in supplementary table S2, Supplementary Material online.

All qPCR experiments were performed in FrameStar 96 well plates (Institute of Applied Biotechnologies [IAB], Prague, Czech Republic) covered by µltraAmp Plate Sealers (Sorenson BioScience, Salt Lake City, UT) or qPCR adhesive foil (IAB) using a C1000 Thermal cycler CFX96 Real-Time System (Bio-Rad, Hercules, CA).

Cloning of Genes of Interest in Other Gelechioid Species

The genes of interest included the reference genes *EF-1a* and *Ace-1*, together with the markers proven to be sex-linked in *T. absoluta*, namely *Pix* and *chitinase h* (*Chit*) for the chromosome homoeologous to *B. betularia* linkage group (BbLG) 7, and *90-kDa heat shock protein* (*Hsp90*) and *twitchin* (*Tw*) for the chromosome homoeologous to BbLG27 (see results for details). Degenerate primers (supplementary table S5, Supplementary Material online) were designed for regions of coding sequences conserved between Lepidoptera and other insect species, and used for RT-PCR amplification of partial sequences with the first-strand cDNA as a template. Amplified fragments were cloned using pGEM-T Easy Vector System (Promega, Madison, WI) or CloneJET PCR cloning kit (Thermo Fisher Scientific, Waltham, MA), and confirmed by Sanger sequencing. The obtained sequences were deposited in GenBank (for accession numbers, see supplementary table S1, Supplementary Material online) and used for the design of species-specific primers for qPCR experiments (supplementary table S4, Supplementary Material online).

Quantitative Analysis of Gene Dose in the Other Gelechioid Species

Quantitative PCR experiments using male and female gDNAs as template were conducted in *S. cerealella*, *P. operculella*, *C. laricella*, *O. arenosella*, *H. pseudospretella*, and *D. daucella* to test for the sex-linkage of *Pix*, *Chit*, *Hsp90*, and *Tw*. Male and female gene doses of the target genes were compared with *EF-1a* and/or *Ace-1*. Three technical and three biological replicates were used per experiment. Composition of the reactions, cycling conditions, and sequences of forward and reverse species-specific primers are detailed in supplementary tables S3 and S4, Supplementary Material online. The F:M ratio was calculated including the *E* value of the primers, according to the formula mentioned earlier, and then compared with the expected values of 1 and 0.5 corresponding to autosomal position and sex-linkage, respectively, by means of one-sample *t*-test using R.

Results

Barcoding of Collected Specimens

The field collected larvae used for chromosome preparations were barcoded using a partial sequence of *COI*. The sequences confirmed the classification of *H. pseudospretella* (Oecophoridae) and *D. daucella* (Depressaridae) with 100% identity with their respective records in the BOLD database. In the case of the Coleophoridae specimens, our search retrieved matches with *C. laricella* and *Coleophora sibiricella*. Since the geographical distribution of both species does not overlap in the Czech Republic (Laštůvka and Liška 2011), we considered the samples as *C. laricella* in our analysis. The consensus sequences of the *COI* fragments for all species examined were deposited in the GenBank database under accession numbers detailed in supplementary table S1, Supplementary Material online.

Karyotype Analyses

In comparison with the most common and ancestral lepidopteran chromosome number $n = 31$ (see Discussion for details), all species of Gelechioidea studied herein showed a reduced chromosome number ranging from $n = 28$ to $n = 30$. These values are in concordance with those observed in other representatives of the superfamily, which shows a modal chromosome number of $n = 29$ in 15 out of 33 studied species (supplementary table S6, Supplementary Material online). FISH with the telomeric probe marking chromosome ends was used to accurately count chromosome numbers in some of the examined species (cf. Šíchová et al. 2015; not shown).

In the Gelechiid assemblage, a complete karyotype analysis including the identification of sex chromosome constitution analysis has not been performed except for *P. operculella* (Gelechiidae) (Bedo 1984; Makee and Tafesh 2006; supplementary table S6, Supplementary Material online). In the present study, we analyzed two representatives of the family Gelechiidae, namely *T. absoluta* and *S. cerealella*.

In *T. absoluta*, Carabajal Paladino et al. (2016) determined the haploid chromosome number of $n = 29$ and identified the largest elements as sex chromosomes morphometrically. In the present study, we identified the W chromosome by

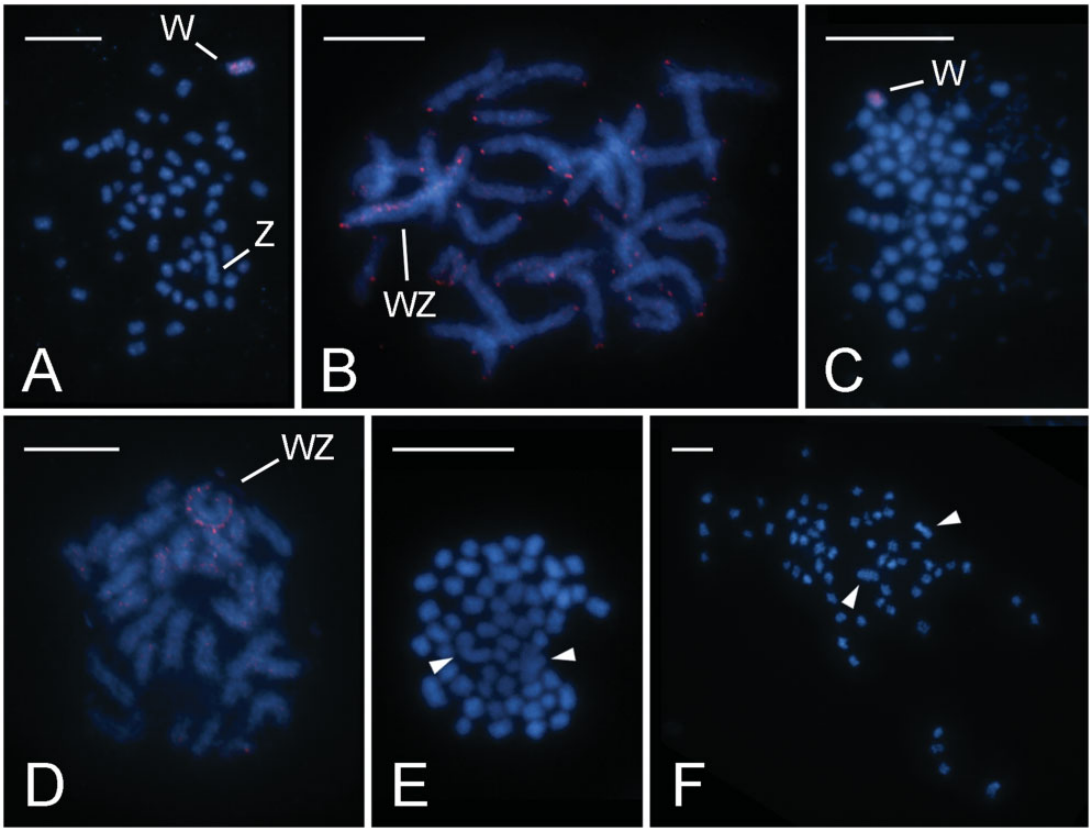

Fig. 1.—Cytogenetic analysis of representatives of the Gelechiid and Scythridid assemblages. Chromosomes were counterstained with DAPI (blue); female derived genomic probes (*A*–*D*) were labeled by Cy3 (red). (*A* and *B*) GISH in *Tuta absoluta* (Gelechiidae, Gelechiid assembl.): (*A*) female mitotic metaphase consisting of $2n = 58$ elements; note that the W chromosome is one of the two largest chromosomes in the complement; (*B*) female pachytene nucleus; the probe labeled the W chromosome in the WZ bivalent and chromosome ends of most bivalents. (*C* and *D*) GISH in *Sitotroga cerealella* (Gelechiidae, Gelechiid assembl.): (*C*) female mitotic metaphase consisting of $2n = 60$ chromosomes; the W chromosome is not conspicuously larger than the other chromosomes; note DAPI-stained small rod-shaped bodies, probably corresponding to bacteria; (*D*) late pachytene female nucleus; the probe identified the W chromosome in the WZ bivalent. (*E* and *F*) Mitotic complements of *Coleophora laricella* (Coleophoridae, Scythridid assembl.) stained with DAPI: (*E*) male mitotic metaphase consisting of $2n = 58$ chromosomes; note a pair of large chromosomes (arrowheads); (*F*) female mitotic metaphase comprising $2n = 58$ chromosomes; note a pair of large chromosomes (arrowheads). Bar = 10 μm.

means of GISH, in which the labeled female gDNA-derived probe was hybridized to chromosomes in excess of unlabeled male competitor DNA. In mitotic complements, hybridization signals clearly highlighted one chromosome of the large pair. GISH thus confirmed that this is the W chromosome and implied that the other large element represents the Z chromosome (fig. 1*A*). The probe produced signals scattered along the W chromosome with notable exception of one subtelomeric and one interstitial gap in pachytene nuclei, and in some experiments also highlighted the chromosome ends (fig. 1*B*).

The haploid chromosome number of $n = 30$ was previously described for males of *S. cerealella* (Lukhtanov and Kuznetsova 1989). We confirmed the chromosome number in mitotic complements, $2n = 60$, in males (not shown) as well as in females (fig. 1*C*). Furthermore, we used GISH to identify the female-specific W chromosome in mitotic complements (fig. 1*C*). In most mitotic metaphases, the W chromosome was not clearly discernible by size. In order to improve the resolution, GISH experiments were performed on female preparations of elongated pachytene bivalents. These experiments provided a more informative labeling pattern of the female genomic probe on the W chromosome. Hybridization signals of the probe were scattered along the entire W chromosome (fig. 1*D*). Interestingly, chromosome preparations obtained from the Argentinian *S. cerealella* females were contaminated with small DAPI-positive bodies, most likely corresponding to some bacteria present in the ovaries (fig. 1*C*).

Coleophora laricella (Coleophoridae) was the only representative of the Scythridid assemblage examined in this study. Mitotic metaphase complements consisted of $n = 29$ in both males and females of this species. The karyotype of both sexes comprised a conspicuously large chromosome pair (fig. 1*E*

and *F*). Surprisingly, GISH provided weak or no hybridization signals in mitotic nuclei (not shown). GISH carried out on less condensed female pachytene chromosomes failed to identify a W-chromosome as well (not shown). Telomeric FISH combined with GISH was used as a control and yielded clear telomeric but no GISH signals (not shown). So it seems that our negative GISH results are not artifactual but rather point to an exceptional molecular composition of the *C. laricella* W chromosome. The W chromosome of *C. laricella* presumably does not differ from the rest of the genome in that it comprises a diverse spectrum of ubiquitous repeats present at low abundance.

Within the Depressariid assemblage, three species, namely *O. arenosella* (Xyloryctidae), *H. pseudospretella* (Oecophoridae), and *D. daucella* (Depressaridae), were investigated. In *O. arenosella*, the diploid chromosome complement consisted of $2n = 60$ chromosomes in both males (fig. 2*A*) and females (fig. 2*B*). No elements showed significant size differences in *O. arenosella*, with all chromosomes decreasing gradually in size, which is typical for lepidopteran karyotypes (fig. 2*A* and *B*). In addition, no mitotic chromosome was reliably discerned by GISH in this species as the female-derived genomic probe labeled all chromosomes more or less with the same intensity (not shown). In pachytene, the probe labeled all bivalents, some along the entire chromosome length and some preferentially in subterminal regions. However, one bivalent was conspicuous by its heteromorphic staining with one of its threads intensively stained while the other was not (fig. 2*C*). It is reasonable to assume that this bivalent corresponds to the WZ sex chromosome pair. The absence of hybridization signals on the Z chromosome is likely a result of its hemizygosity in females from which the GISH probe was derived. The sex chromosome bivalent identity was further supported by its meiotic pairing pattern, as the signal-free chromosome typically twisted several times around its labeled partner (fig. 2*C*). This was due to the size difference between the sex chromosomes with the W being much shorter than the Z chromosome (cf. Marec and Traut 1994).

In *H. pseudospretella*, a reduced diploid chromosome number of $2n = 56$ with two large chromosomes was observed in mitotic metaphase nuclei of both sexes (fig. 2*D* and *E*). The female-derived genomic probe clearly highlighted one of the large chromosomes in female mitotic metaphase complements (fig. 2*E*). Thus, the largest chromosome pair most likely comprises the sex chromosomes. However, in female pachytene nuclei, a WZ bivalent could not be identified without the use of GISH. This method revealed a bipartite organization of the W chromosome, as it strongly labeled one terminal region corresponding to roughly one-third of the sex chromosome bivalent (fig. 2*F*).

The diploid chromosome number was $2n = 60$ in both sexes of *D. daucella*. Neither male nor female mitotic complement comprised any notably larger chromosome (fig. 2*G* and *H*). GISH identified one of the larger chromosomes as the W chromosome in the *D. daucella* female mitotic metaphase complements (fig. 2*H*). In female pachytene nuclei, the WZ bivalent was easily discerned by the heterochromatic W thread (not shown). GISH showed scattered hybridization signals colocalizing with DAPI positive blocks on the W chromosome (fig. 2*I*).

Identification of Sex-Linked Synteny Blocks in Gelechioidea

To identify sex-linked synteny blocks, the sex-linkage of *T. absoluta* genes was tested by qPCR using male and female gDNA as template. This method can detect hemizygosity of Z-linked markers caused either by the absence or molecular degradation of their W-linked gene copies (Nguyen et al. 2013; Dalíková et al. 2017). The variable female-to-male (F:M) ratio between the selected reference genes *EF-1a* and *Ace-1*, using *Ace-1* as target and *EF-1a* as reference, was 1.000 ± 0.102 (SE), which statistically differed from 0.5 ($P < 0.05$) but not from 1 ($P > 0.05$) (supplementary table S3, Supplementary Material online). The F:M ratio for *ket*, using *Ace-1* as reference, gave a value of 0.498 ± 0.090, which significantly differed from 1 ($P < 0.05$) but not from 0.5 ($P > 0.05$) (supplementary table S3, Supplementary Material online). These results indicated that females and males had the same copy number of both *Ace-1* and *EF-1a* genes, and that females had half the number of copies of *ket* with respect to males, which was expected as this gene represents a standard marker for the lepidopteran Z chromosome (cf. Nguyen et al. 2013; Van't Hof 2013). The analysis thus confirmed that the *Ace-1* and *EF-1a* genes are autosomal and can be used as reference genes for further studies. It also proved *ket* as a good control gene for the screening of sex-linked markers in *T. absoluta*.

The results of the screening of marker genes in *T. absoluta* are presented in supplementary table S2, Supplementary Material online and figure 3. Markers orthologous to genes of *B. betularia* (Geometridae) LG1 (*ket*), LG7 (*Pix*) and LG27 (*Hsp90*) were sex-linked in this species, with F:M ratios ranging from 0.491 (*ket*) to 0.590 (*Hsp90*), considering the values obtained with both reference genes (*EF-1a* and *Ace-1*). The rest of the markers ranged from 0.800 for *ribosomal protein L4* (marker for BbLG29) to 1.508 for *18–56 protein* (marker for BbLG20), and were considered autosomal. Deviation of markers from the expected F:M value of 1 could be attributed to differences in primer efficiency, which was not corrected in the initial screening.

BbLG1 corresponds to the Z chromosomes in the ancestral karyotype of $n = 31$, while the other two chromosomes (BbLG7 and BbLG27) are autosomes. An extra marker gene was hence considered for further analysis of these autosomes: *Chit* for BbLG7 and *Tw* for BbLG27. Orthologs of all four marker and both reference genes were then amplified and cloned from *P. operculella*, *S. cerealella*, *C. laricella*,

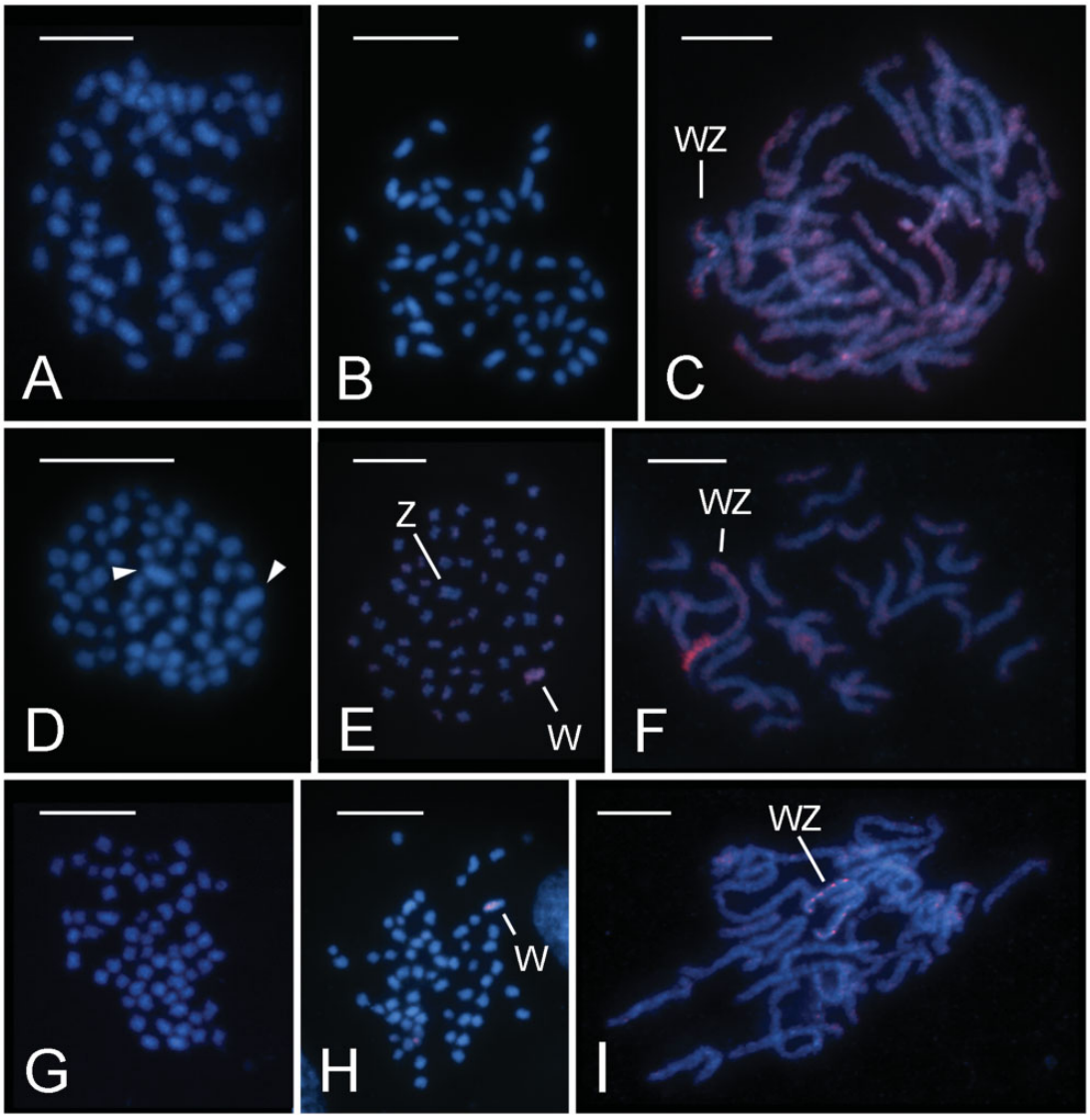

Fig. 2.—Cytogenetic analysis of representative of the Depressariid assemblage. Chromosomes were counterstained with DAPI (blue); female-derived genomic probes (*C*, *E*, *F*, *H*, *I*) were labeled by Cy3 (red). (*A*–*C*) *Opisina arenosella* (Xyloryctidae): (*A*) male mitotic metaphase consisting of $2n = 60$ elements; (*B*) female mitotic metaphase consisting of $2n = 60$ chromosomes; (*C*) GISH on female pachytene nucleus; note the hybridization signals on all bivalents either along the entire chromosomes or with preference for subterminal regions; the WZ bivalent is identified by the signal intensity that differs between the W and Z chromosome threads, as well as by the characteristic pairing of the longer Z chromosome twisted around the much shorter W chromosome. (*D*–*F*) *Hofmannophila pseudospretella* (Oecophoridae): (*D*) male mitotic metaphase consisting of $2n = 56$ chromosomes; note the two largest chromosomes (arrowheads); (*E* and *F*) GISH on female chromosome preparations; (*E*) female mitotic metaphase comprising $2n = 56$ chromosomes; note that GISH identified the W chromosome as one of the two largest chromosomes; (*F*) female pachytene nucleus; note the size and bipartite organization of the WZ bivalent with about one-third of the W chromosome thread strongly labeled with the probe. (*G*–*I*) *Depressaria daucella* (Depressariidae): (*G*) male mitotic metaphase comprising $2n = 60$ chromosomes; note that there is no conspicuously larger chromosome pair; (*G*–*I*) GISH on female chromosome preparations; (*H*) female mitotic metaphase consisting of $2n = 60$ elements with the W chromosome identified by the probe; (*I*) female pachytene nucleus; note the WZ bivalent showing scattered hybridization signals of the probe on the W chromosome thread. Bar = 10 μm.

O. arenosella, *H. pseudospretella*, and *D. daucella*. The partial sequences were deposited in NCBI (accession numbers in supplementary table S1, Supplementary Material online) and used for the design of species-specific primers for qPCR experiments (supplementary table S4, Supplementary Material online).

The results in *T. absoluta* and the other gelechioid species are shown in supplementary table S3, Supplementary Material online and summarized in figure 4. The F:M ratio values for both chromosomal markers corresponding to BbLG7 significantly differed from 1 ($P < 0.05$) but not from 0.5 ($P > 0.05$) in all species except for *D. daucella*, which

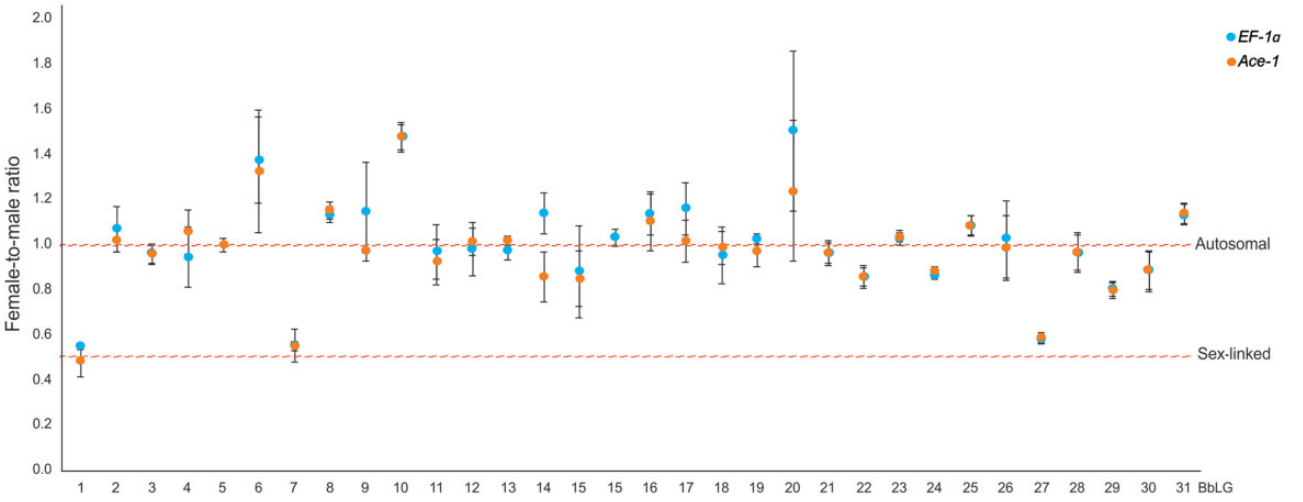

FIG. 3.—Screening of marker genes in *Tuta absoluta* by means of qPCR. Blue dots represent the average female-to-male ratio values obtained for each marker using *EF-1a* as the reference gene. Orange dots are the average values for the same variable obtained using *Ace-1* as the reference gene. Whiskers show the SE. Red dashed lines are used to show how each value correlates with 1 (autosomal) and 0.5 (sex-linked) expected female-to-male ratios. Note that most of the data points fluctuate ~1, except for those corresponding to BbLG1, BbLG7, and BbLG27 which are closer to 0.5 than to 1. BbLG, *Biston betularia* linkage group.

suggested that the markers were sex-linked. In *D. daucella*, the F:M ratio of *Chit* was 0. 528 ± 0.021, while for *Pix* it was 0. 861 ± 0.074 (*EF-1a* as the reference gene), which is consistent with sex-linkage of the former and autosomal inheritance of the latter.

For the BbLG27 markers, the F:M ratio statistically differed from 0.5 ($P < 0.05$) but not from 1 ($P > 0.05$) in *S. cerealella*, *C. laricella*, *O. arenosella*, *H. pseudospretella*, and *D. daucella*, indicating that the markers had an autosomal location. The opposite situation was observed in *T. absoluta*, meaning that both markers were sex-linked in this species. Interesting results were obtained in *P. operculella*, where *Tw* was sex-linked but *Hsp90* was not (F:M ratios of 0. 578 ± 0. 019 and 0.999 ± 0.047, respectively; *EF-1a* as the reference gene). These findings, together with the discrepancies found for the markers for BbLG7 in *D. daucella*, were corroborated using the second reference gene (*Ace-1*) with a similar outcome (supplementary table S3, Supplementary Material online).

Discussion

In this study, we analyzed the sex chromosomes of seven species sampled across all three major lineages of the superfamily Gelechioidea (cf. Sohn et al. 2016; for phylogenetic relationships, see fig. 4 and supplementary fig. S1, Supplementary Material online). All species under study have a derived chromosome number compared with the ancestral lepidopteran karyotype of $n = 31$. Our cytogenetic analyses confirmed the expected presence of a large chromosome pair in the karyotypes of *T. absoluta* (Gelechiidae), *C. laricella* (Coleophoridae), and *H. pseudospretella* (Oecophoridae), species with karyotypes reduced to $n = 29$ in the first two and $n = 28$ in the latter (figs. 1*A, B, E, F* and 2*D, E*). The existence of a conspicuously large chromosome pair was a characteristic feature of the Gelechioidea karyotypes described to date (supplementary table S6, Supplementary Material online) and Ennis (1976) regarded them as autosomal fusion products. The GISH experiments performed in this study, however, confirmed that the largest chromosome pairs are indeed sex chromosomes in *T. absoluta* and *H. pseudospretella* (figs. 1*A* and 2*E*). In *C. laricella*, the W chromosome could not be identified (not shown). Thus, our cytogenetic data suggest that the largest chromosome pair corresponds to sex chromosomes only in some gelechioid species. A similar size difference, that is, the largest chromosome pair being about 1.5–2 times larger than the second largest one in a descending size series, was also observed in other *Coleophora* species (Lukhtanov and Puplesiene 1999) and in *P. operculella* (Gelechiidae) (Bedo 1984) suggesting chromosome fusions. Interspecific differences were observed in the relative size of the sex chromosomes, which were not so conspicuous in species with $n = 30$, namely *S. cerealella* (Gelechiidae), *O. arenosella* (Xyloryctidae), and *D. daucella* (Depressaridae) (figs. 1*C, D* and 2*A–C, G–I*). A larger chromosome pair, which was not detected in our study, was reported for *S. cerealella* by Lukhtanov and Kuznetsova (1989) based on preparations of metaphase I bivalents from males (supplementary table S6, Supplementary Material online). This inconsistency could be caused by different methods, tissues used for chromosome preparations, and the type of cell division.

To confirm the fusions and identify the synteny blocks involved, we tested selected markers for all chromosomes of the ancestral karyotype with $n = 31$ (Van't Hof 2013; Ahola et al. 2014) for their sex-linkage in *T. absoluta* by means of

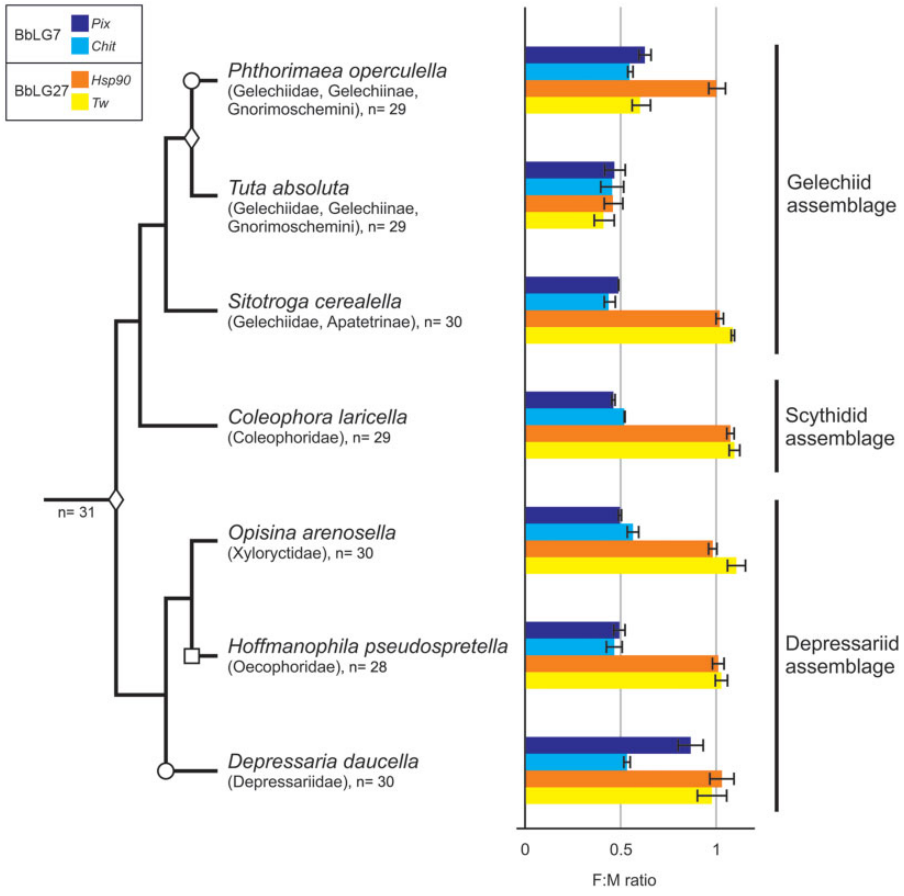

Fig. 4.—Phylogenetic relationship between the species analyzed in this study, including a graphic representation of the results obtained using qPCR for the analysis of selected marker genes. Bar charts show the obtained female-to-male ratios (including SEs) of the copy number of the selected marker genes *Pix* and *Chit* for BbLG7, and *Hsp90* and *Tw* for BbLG27, using *EF-1a* as the reference gene. Values close to 0.5 indicate sex-linkage, while values close to 1 indicate autosomal location of the marker. F:M ratio, female-to-male ratio. Note that the decrease in chromosome numbers coincides with sex chromosome–autosome fusions confirmed by qPCR. Diamond, confirmed fusion; circle, translocation or incomplete degeneration of one marker; square, putative fusion suggested by cytogenetic data.

qPCR. The qPCR results confirmed the sex-linkage of markers located on the Z chromosome in other Lepidoptera (Nguyen et al. 2013; Van't Hof 2013) and identified synteny blocks homoeologous to *B. betularia* (Geometridae) linkage group (BbLG) 7 and 27 as candidates for fusions (fig. 3). Testing of two markers for each chromosome, namely *Pix* and *Chit* for BbLG7, and *Hsp90* and *Tw* for BbLG27, confirmed their sex-linkage in *T. absoluta* and thus strongly supported fusions of these synteny blocks with the ancestral Z chromosome (supplementary table S3, Supplementary Material online and fig. 4). qPCR analyses of *Pix* and *Chit* in the other species showed a sex-linkage of both markers in all gelechioids but *D. daucella*, in which only *Chit* and not *Pix* was sex-linked (supplementary table S3, Supplementary Material online and fig. 4). Assuming current phylogenetic hypotheses (Heikkilä et al. 2014; Sohn et al. 2016; supplementary fig. S1, Supplementary Material online), the qPCR results suggest that the fusion of the Z chromosome and chromosome homoeologous to BbLG7 [hereinafter F(Z; 7)] occurred in a common ancestor of the superfamily Gelechioidea. Thus, the autosomal location of *Pix* in *D. daucella* most likely points to a secondary translocation of this gene to an unidentified autosome (cf. Nguyen et al. 2013) or the W chromosome (Van't Hof 2013) or to incomplete degeneration of its W-linked copy. The latter, however, seems unlikely in this case, as the F(Z; 7) fusion occurred ~100 Ma (Wahlberg et al. 2013). Sex-linkage analyses of *Hsp90* and *Tw* revealed that these markers are autosomal in all species but two representatives of the family Gelechiidae, *T. absoluta* and *P. operculella*, with *Tw* sex-linked in the latter but not *Hsp90* (supplementary table S3, Supplementary Material online and fig. 4). This, together with the autosomal localization of both markers in *S. cerealella*, suggests that the neo-Z chromosome formed by the F(Z; 7) fusion further fused with BbLG27 [hereinafter F(neo-Z; 27)] in a common ancestor of the tribe Gnorimoschemini. However, we cannot exclude the possibility that the F(neo-

Z; 27) fusion occurred earlier in the subfamily Gelechiinae (cf. Karsholt et al. 2013). Autosomal linkage of *Hsp90* in *P. opercullella* can again be explained by its translocation (see above). However, given the relative young age of the F(neo-Z; 27) fusion, we cannot fully rule out the *Hsp90* allele persisting on the neo-W chromosome. Further research is needed to trace the exact evolutionary origin and level of differentiation of the F(neo-Z; 27) fusion. Moreover, the reduced chromosome number observed in *H. pseudospretella* (see above), the large size of its neo-sex chromosome pair along with the partial differentiation of its W chromosome suggest that another fusion between the F(Z; 7) and an autosome occurred independently in the family Oecophoridae.

Our results hence clearly show that at least two sex chromosome–autosome fusions occurred in the evolution of the diverse superfamily Gelechioidea. This finding further adds to the growing list of derived sex chromosome systems recently identified in various lepidopteran taxa, such as leafrollers of the family Tortricidae (Nguyen et al. 2013; Šíchová et al. 2013; Picq et al. 2018), leaf miners of the family Gracillaridae (Dalíková et al. 2017; Fraïsse et al. 2017), and *Leptidea* wood white (Pieridae) (Šíchová et al. 2015, 2016) and *Danaus* (Nymphalidae) butterflies (Smith et al. 2016; Mongue et al. 2017; Traut et al. 2017). The latter represent yet another case of repeated sex chromosome–autosome fusions, similar to those reported in this study. All these findings illustrate that neo-sex chromosomes are not exceptional in moths and butterflies. Rather, they appear to be relatively common, not only in terms of number of species, as the Tortricidae and Gelechioidea taxa alone comprise together about 17% of the described lepidopteran biodiversity (Beccaloni et al. 2018) but also in the number of independent origins (Nguyen and Carabajal Paladino 2016; cf. Pokorná et al. 2014). This suggests that the paucity of sex chromosome–autosome fusions is not an intrinsic feature of female heterogamety as previously assumed (Pokorná et al. 2014; Pennell et al. 2015).

Lepidoptera possess holokinetic chromosomes, which attach to kinetochore microtubules along most of the chromosomal surface (Wolf 1994). This reduces the risk of formation of dicentric and acentric chromosomes and hence it is expected to facilitate chromosomal rearrangements (Wrensch et al. 1994). Indeed, high variation in chromosome numbers was observed in moths and butterflies (Blackmon et al. 2017). However, this genome instability is confined only to a few lepidopteran taxa (Robinson 1971; Talavera et al. 2013). Comparative genomic studies have revealed that lepidopteran karyotypes are very stable with the modal chromosome number of $n = 31$ being the ancestral one. Furthermore, it has been shown that chromosome fusions are not random in this insect order since independent fusions observed in distant species involve the same small and repeat-rich chromosomes (Van't Hof 2013; Ahola et al. 2014). Reconstructions of karyotype evolution in several lepidopteran clades with derived sex chromosome systems also show that the first large-scale chromosome rearrangements which differentiated the karyotypes of examined taxa from the ancestral $n = 31$ tend to be sex chromosome–autosome fusions (Nilsson et al. 2008; Nguyen et al. 2013; Šíchová et al. 2013; Dalíková et al. 2017; Mongue et al. 2017). Although the reconstruction of karyotype evolution in a group so diverse as Gelechioidea is challenging due to the scarcity of available data (supplementary table S6, Supplementary Material online), the reduced chromosome number of $n = 30$ in families Gelechiidae, Elachistidae, Xyloryctidae, and Depressariidae suggests that the F(Z; 7) fusion occurred early in the karyotype evolution of gelechioids.

This propensity of lepidopteran sex chromosomes for fusions could shed light on the evolutionary forces driving chromosomal change. The higher rate of sex chromosome–autosome fusions in XX/XY than in WZ/ZZ systems observed in vertebrates (Pokorná et al. 2014; Pennell et al. 2015) led to the conclusion that fusions must be driven by two or more evolutionary forces (Pennell et al. 2015; Kirkpatrick 2017). A simpler explanation for the higher rate of Y-autosome fusions in vertebrates, random genetic drift (Kirkpatrick 2017), was dismissed due to the lack of multiple sex chromosomes in female heterogametic groups (Pennell et al. 2015; Kirkpatrick 2017). Genetic drift, however, can be invoked to explain the high incidence of neo-sex chromosomes in Lepidoptera. In such case, the same pattern observed in vertebrates (a higher incidence of W–autosome than Z–autosome fusions) is expected for lepidopteran multiple sex chromosome systems. However, the W–autosome and Z–autosome fusions resulting in multiple sex chromosome constitutions WZ_1Z_2 and W_1W_2Z, respectively, observed so far in Lepidoptera are tied (Traut et al. 2007; Šíchová et al. 2015, 2016; Smith et al. 2016). Furthermore, many of the other recently reported neo-sex chromosomes systems are not informative as males and females exhibit the same chromosome number (Nguyen et al. 2013; Dalíková et al. 2017; Fraïsse et al. 2017; Mongue et al. 2017; this study). Available data thus do not allow us to evaluate the role of genetic drift in sex chromosome–autosome fusions in Lepidoptera.

Chromosome rearrangements such as fusions or inversions affect linkage relationships and thus can play an important role in adaptation and speciation (Yeaman 2013; Charlesworth 2015; Ortiz-Barrientos et al. 2016). In leafrollers of the family Tortricidae, Nguyen et al. (2013) reported the fusion of the Z chromosome with an autosome homoeologous to BbLG15. This chromosome is enriched in genes involved in detoxification and regulated absorption of plant secondary metabolites, namely esterases and ABC transporters, which are crucial for the performance of lepidopteran larvae on their host plants. The fusion thus linked these performance genes together with sex-linked female preference or host-independent isolation genes, which can facilitate

adaptation and speciation in the presence of gene flow (Matsubayashi et al. 2010). Furthermore, it was hypothesized that the neo-Z-linked performance genes got amplified to make up for their nonrecombining and thus gradually degenerating maternally inherited gametologues (Nguyen et al. 2013). Following functional divergence of the new performance gene copies supposedly contributed to adaptation to new hosts which could eventually result in the formation of new species (cf. Li et al. 2003). Interestingly, BbLG7, which is involved in the F(Z; 7) fusion shared by all gelechioids, comprises the largest cluster of UDP-glycosyltransferases (UGTs). Enzymes encoded by the UGT gene family catalyze the glycosylation of small lipophilic compounds, turning them into water-soluble and thus more easily excreted products (Ahn et al. 2012). Although UGTs have been considerably understudied compared with other detoxification families, evidence supporting their role in detoxification of plant secondary metabolites and insecticides in Lepidoptera has been growing (Ahn et al. 2011; Wouters et al. 2014; Krempl et al. 2016; Li et al. 2017). Therefore, we hypothesize that the sex chromosome–autosome fusions may indeed contribute to ecological specialization and speciation in moths.

Sex chromosome turnover has been shown to predate, so far, two large lepidopteran radiations, Tortricidae and Gelechioidea (Nguyen et al. 2013; this study). The F(Z; 7) fusion observed in gelechioids fits well the scenario drawn by Nguyen et al. (2013) and the enrichment in performance genes of the autosomes involved in fusions in both lineages points to more general aspects of the lepidopteran karyotype evolution. The superfamily Gelechioidea provides an opportunity to test the hypothesis on the role of neo-sex chromosomes in the speciation of Lepidoptera, as sister lineages with and without neo-sex chromosomes of different age can be examined in parallel, along with their diversification rates.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Our thanks go to Hayat Makee and Noha Tafesh (Damascus, Syria) for cultures of *P. operculella*, Cornel Adler (Berlin, Germany) for cultures of *S. cerealella*, Cynthia L. Cagnotti (Castelar, Argentina) for cultures of *T. absoluta*, and Marie Korchová for rearing insects and technical assistance. This work was supported by the Czech Science Foundation (14-35819 P and 17-17211S to P.N. and 17-13713S to F.M.) and the project Postdok_BIOGLOBE (CZ.1.07/2.3.00/30.0032) cofinanced by the European Social Fund and the state budget of the Czech Republic.

Literature Cited

Ahn S-J, et al. 2011. Metabolic detoxification of capsaicin by UDP-glycosyltransferase in three *Helicoverpa* species. Arch Insect Biochem Physiol. 78(2):104–118.

Ahn SJ, Vogel H, Heckel DG. 2012. Comparative analysis of the UDP-glycosyltransferase multigene family in insects. Insect Biochem Mol Biol. 42(2):133–147.

Ahola V, et al. 2014. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nat Commun. 5:4737.

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet. 14(2):113–124.

Beccaloni G, et al. 2018. LepIndex: the global Lepidoptera names index. In: Roskov Y, et al., editors. Species 2000 & ITIS catalogue of life, 2018 annual checklist. Leiden: Species 2000: Naturalis.

Bedo DG. 1984. Karyotypic and chromosome-banding studies of the potato tuber moth, *Phthorimaea operculella* (Zeller) (Lepidoptera, Gelechiidae). Can J Genet Cytol. 26(2):141–145.

Berger M, et al. 2016. Insecticide resistance mediated by an exon skipping event. Mol Ecol. 25(22):5692–5704.

Blackmon H, Ross L, Bachtrog D. 2017. Sex determination, sex chromosomes, and karyotype evolution in insects. J Hered. 108(1):78–93.

Cagnotti CL, et al. 2012. Effects of X-rays on *Tuta absoluta* for use in inherited sterility programmes. J Pest Sci. 85(4):413–421.

Carabajal Paladino LZ, et al. 2016. The effect of X-rays on cytological traits of *Tuta absoluta* (Lepidoptera: Gelechiidae). Fla Entomol. 99(sp1):43–53.

Charlesworth D. 2015. Plant contributions to our understanding of sex chromosome evolution. New Phytol. 208(1):52–65.

Dalíková M, et al. 2017. New insights into the evolution of the W chromosome in Lepidoptera. J Hered. 108(7):709–719.

Delph LF, Demuth JP. 2016. Haldane's Rule: genetic bases and their empirical support. J Hered. 107(5):383–391.

Dufresnes C, et al. 2016. Empirical evidence for large X-effects in animals with undifferentiated sex chromosomes. Sci Rep. 6:21029.

Ennis TJ. 1976. Sex chromatin and chromosome numbers in Lepidoptera. Can J Genet Cytol. 18(1):119–130.

Fraïsse C, Picard MAL, Vicoso B. 2017. The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. Nat Commun. 8(1):1486.

Graves J. 2016. Did sex chromosome turnover promote divergence of the major mammal groups? Bioessays 38(8):734–743.

Haldane J. 1922. Sex ratio and unisexual sterility in hybrid animals. J Gen. 12(2):101–109.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci U S A. 1101:4812–14817.

Heikkilä M, Mutanen M, Kekkonen M, Kaila L. 2014. Morphology reinforces proposed molecular phylogenetic affinities: a revised classification for Gelechioidea (Lepidoptera). Cladistics 30(6):563–589.

Karsholt O, Mutanen M, Lee S, Kaila L. 2013. A molecular analysis of the Gelechiidae (Lepidoptera, Gelechioidea) with an interpretative grouping of its taxa. Syst Entomol. 38(2):334–348.

Kato A, Albert P, Vega J, Birchler J. 2006. Sensitive fluorescence in situ hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. Biotech Histochem. 81(2–3):71–78.

Kirkpatrick M. 2017. The evolution of genome structure by natural and sexual selection. J Hered. 108(1):3–11.

Kitano J, Peichel CL. 2012. Turnover of sex chromosomes and speciation in fishes. Environ Biol Fishes. 94(3):549–558.

Kitano J, et al. 2009. A role for a neo-sex chromosome in stickleback speciation. Nature 461(7267):1079–1083.
Krempl C, et al. 2016. Potential detoxification of gossypol by UDP-glycosyltransferases in the two Heliothine moth species *Helicoverpa armigera* and *Heliothis virescens*. Insect Biochem Mol Biol. 71:49–57.
Laštůvka Z, Liška J. 2011. Annotated checklist of moths and butterflies of the Czech Republic (Insecta: Lepidoptera). Brno: Biocont Laboratory.
Li WM, Schuler MA, Berenbaum MR. 2003. Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: specificity and substrate encounter rate. Proc Natl Acad Sci U S A. 100(Suppl 2):14593–14598.
Li X, Zhu B, Gao X, Liang P. 2017. Over-expression of UDP-glycosyltransferase gene *UGT2B17* is involved in chlorantraniliprole resistance in *Plutella xylostella* (L.). Pest Manag Sci. 73(7):1402–1409.
Lima TG. 2014. Higher levels of sex chromosome heteromorphism are associated with markedly stronger reproductive isolation. Nat Commun. 5:4743.
Lukhtanov VA, Kuznetsova VG. 1989. Karyotype structure in higher Lepidoptera (Papilionomorpha.). Entomol Rev. 68:12–31.
Lukhtanov VA, Puplesiene J. 1999. Polyploidy in bisexual Lepidoptera species (Insecta: Lepidoptera): old hypotheses and new data. Bonn Zool Beitr. 48:313–328.
Makee H, Tafesh N. 2006. Sex chromatin body as a marker of radiation-induced sex chromosome aberrations in the potato tuber moth, *Phthorimaea operculella* (Lepidoptera: Gelechiidae). J Pest Sci. 79(2):75–82.
Mank JE, Hosken DJ, Wedell N. 2014. Conflict on the sex chromosomes: cause, effect, and complexity. Cold Spring Harb Perspect Biol. 6(12):a017715.
Marec F, Traut W. 1994. Sex chromosome pairing and sex chromatin bodies in W-Z translocation strains of *Ephestia kuehniella* (Lepidoptera). Genome 37(3):426–435.
Masly JP, Presgraves DC. 2007. High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. PLoS Biol. 5:1890–1898.
Matsubayashi KW, Ohshima I, Nosil P. 2010. Ecological speciation in phytophagous insects. Entomol Exp Appl. 134(1):1–27.
Méndez LM, et al. 2016. Evaluation of three wheat types for the rearing of *Sitotroga cerealella* Olivier (Lepidoptera: Gelechiidae) commonly used on natural enemies rearing. Rev Soc Entomol Argent. 75:105–116.
Mongue AJ, Nguyen P, Voleníková A, Walters JR. 2017. Neo-sex chromosomes in the Monarch butterfly, *Danaus plexippus*. G3 7(10):3281–3294.
Natri HM, Shikano T, Merila J. 2013. Progressive recombination suppression and differentiation in recently evolved neo-sex chromosomes. Mol Biol Evol. 30(5):1131–1144.
Nguyen P, Carabajal Paladino LZ. 2016. Evolutionary biology: convergent evolution, evolution of complex traits, concepts and methods In: Pontarotti P, editor. On the neo-sex chromosomes of Lepidoptera. Heidelberg: Springer. p. 171–185.
Nguyen P, et al. 2013. Neo-sex chromosomes and adaptive potential in tortricid pests. Proc Natl Acad Sci U S A. 110(17):6931–6936.
Nilsson NO, Löfstedt C, Dävring L. 2008. Unusual sex-chromosome inheritance in six species of small ermine moths (*Yponomeuta*, Yponomeutidae, Lepidoptera). Hereditas 108(2):259–265.
Ortiz-Barrientos D, Engelstädter J, Rieseberg LH. 2016. Recombination rate evolution and the origin of species. Trends Ecol Evol. 31(3):226–236.
Pala I, Hasselquist D, Bensch S, Hansson B. 2012. Patterns of molecular evolution of an avian neo-sex chromosome. Mol Biol Evol. 29(12):3741–3754.
Pennell MW, et al. 2015. Y fuse? Sex chromosome fusions in fishes and reptiles. PLoS Genet. 11(5):e1005237.
Picq S. et al. 2018. Insights into the structure of the spruce budworm (*Choristoneura fumiferana*) genome, as revealed by molecular cytogenetic analyses and a high-density linkage map. G3 8:2539–2549.
Pokorná M, Altmanová M, Kratochvíl L. 2014. Multiple sex chromosomes in the light of female meiotic drive in amniote vertebrates. Chromosome Res. 22(1):35–44.
R Core Team. 2013. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available from: http://www.R-project.org/; last accessed April 16, 2019.
Ratnasingham S, Hebert P. 2007. BOLD: the Barcode of Life Data System. Mol Ecol Notes. 7(3):355–364.
Robinson R. 1971. Lepidoptera genetics. Oxford: Pergamon Press.
Sahara K, Marec F, Traut W. 1999. TTAGG telomeric repeats in chromosomes of some insects and other arthropods. Chromosome Res. 7(6):449–460.
Saour G, Makee H. 1997. Radiation induced sterility in male potato tuber moth *Phthorimaea operculella* Zeller (Lep., Gelechiidae). J Appl Entomol. 121(1–5):411–415.
Šíchová J, Nguyen P, Dalíková M, Marec F. 2013. Chromosomal evolution in tortricid moths: conserved karyotypes with diverged features. PLoS One 8(5):e64520.
Šíchová J, et al. 2016. Fissions, fusions, and translocations shaped the karyotype and multiple sex chromosome constitution of the northeast-Asian wood white butterfly, *Leptidea amurensis*. Biol J Linn Soc. 118(3):457–471.
Šíchová J, et al. 2015. Dynamic karyotype evolution and unique sex determination systems in *Leptidea* wood white butterflies. BMC Evol Biol. 15:89.
Smith DAS, et al. 2016. A neo-W chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. Proc R Soc Lond B Biol Sci. 283(1835):20160821.
Sohn J-C, et al. 2016. Phylogeny and feeding trait evolution of the mega-diverse Gelechioidea (Lepidoptera: Obtectomera): new insight from 19 nuclear genes. Syst Entomol. 41(1):112–132.
Talavera G, Lukhtanov VA, Rieppel L, Pierce NE, Vila R. 2013. In the shadow of phylogenetic uncertainty: the recent diversification of *Lysandra* butterflies through chromosomal change. Mol Phyl Evol. 69(3):469–478.
Traut W. 1976. Pachytene mapping in female silkworm, *Bombyx mori* L. (Lepidoptera). Chromosoma 58(3):275–284.
Traut W, Ahola V, Smith DAS. 2017. Karyotypes versus genomes: the nymphalid butterflies *Melitaea cinxia*, *Danaus plexippus*, and *D. chrysippus*. Cytogenet Genome Res. 153(1):46–53.
Traut W, Sahara K, Marec F. 2007. Sex chromosomes and sex determination in Lepidoptera. Sex Dev. 1(6):332–346.
Traut W, Sahara K, Otto TD, Marec F. 1999. Molecular differentiation of sex chromosomes probed by comparative genomic hybridization. Chromosoma 108(3):173–180.
Turelli M, Begun DJ. 1997. Haldanes' rule and X-chromosome size in *Drosophila*. Genetics 147(4):1799–1815.
van Nieukerken EJ, et al. 2011. Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. Zootaxa 3148(1):212–221.
Van't Hof AE. 2013. Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a model of industrial melanism. Heredity 110:283–295.
Wahlberg N, Wheat CW, Peña C. 2013. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLoS One 8(11):e80875.

Wolf KW. 1994. The unique structure of lepidopteran spindles. Int Rev Cytol. 152:1–48.
Wouters FC, et al. 2014. Reglucosylation of the benzoxazinoid DIMBOA with inversion of stereochemical configuration is a detoxification strategy in lepidopteran herbivores. Angew Chem Int Ed Engl. 53(42):11320–11324.
Wrensch DL, Kethley JB, Norton RA. 1994. Mites: ecological and evolutionary analyses of life-history patterns. In: Houck MA, editor. Cytogenetics of holokinetic chromosomes and inverted meiosis: keys to the evolutionary success of mites, with generalizations on eukaryotes. New York: Chapman and Hall. p. 282–343.
Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc Natl Acad Sci U S A. 110(19):E1743–E1751.
Yoshido A, Marec F, Sahara K. 2005. Resolution of sex chromosome constitution by genomic in situ hybridization and fluorescence in situ hybridization with (TTAGG)*n* telomeric probe in some species of Lepidoptera. Chromosoma 114(3):193–202.

Associate editor: Marta Barluenga

3.4. Chapter IV

Multiple sex chromosomes of *Yponomeuta* ermine moths suggest a role of sexual antagonism in sex chromosome turnover in Lepidoptera

Provazníková Irena[1,2], Dalíková Martina[1,3,] Voleníková Anna[1], Roessingh Peter[4], Sahara Ken[5], Provazník Jan[1,2], Šlajsová Miroslava [1], Marec František[6] and Nguyen Petr[1,6,*]

Unpublished results

1 Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic

2 Current address: European Molecular Biology Laboratory, Heidelberg, Germany

3 University of Kansas, Lawrence, Kansas, USA

4 University of Amsterdam, Amsterdam, Netherlands

5 Iwate University, Morioka, Japan

6 Institute of Entomology Biology Centre CAS, Ceske Budejovice, Czech Republic

* Correspondence to:
Petr Nguyen
petr.nguyen@prf.jcu.cz

Abstract

Sex chromosome-autosome fusions give rise to neo-sex chromosomes, which provide an insight into early evolution of sex chromosomes and drivers of chromosomal fusions. While sex chromosome-autosome fusions are scarce in vertebrates with female heterogamety (♀ZW/♂ZZ), they are common in moths and butterflies (Lepidoptera), the most species rich group with heterogametic females. This contradicts theoretical model that assumes chromosome fusions to be random and predicts them to be rare in taxa with high chromosome number such as Lepidoptera. In the present study we analyzed sex chromosomes in nine ermine moths of the genus *Yponomeuta* (Yponomeutidae) and their two outgroups, *Teinoptila gutella* (Yponomeutidae) and *Plutella xylostella* (Plutellidae). We employed genomic *in situ* hybridization to identify sex chromosomes and used a custom designed microarray to identify Z-linked genes. Our results confirmed a multiple sex chromosome system Z_1Z_2W to be present in *T. gutella* and all *Yponomeuta* spp. except for *Y. tokyonella*. The multiple sex chromosome system resulted from a fusion between the W chromosome and autosome homeologous to the *Bombyx mori* chromosome 2 (BmChr2). The BmChr2 bears a cluster of genes with ovary-specific expression which suggests that sexually antagonistic selection could have driven fixation of the fusion in a common ancestor of *Yponomeuta* and *Teinoptila* genera. We hypothesize that sex chromosome turnover in Lepidoptera could be driven by sexual antagonism.

Introduction

Sex chromosomes represent a specific part of genome. They are subject to a selection regime distinct from autosomes and play an important role in evolution (Payseur et al. 2018; Connallon et al. 2018). The canonical model of sex chromosome evolution postulates that sex chromosomes arise from a pair of autosomes. When this pair acquires a sex determining (SD) locus, one of its alleles is limited to a heterogametic sex (XY males or ZW females). Selection should favor a linkage disequilibrium between the sex-limited allele and sexually antagonistic (SA) mutations, i.e. mutations beneficial to the heterogametic but detrimental to the homogametic sex (XX females or ZZ males). Suppression of recombination between SD and SA loci is then advantageous and results in differentiation of the sex chromosome pair via accumulation of repeats and deleterious mutations due to Hill-Robertson interactions (Wright et al. 2016; Kratochvíl et al. 2021).

The SA selection should also drive evolution of neo-sex chromosomes resulting from sex chromosome-autosome fusions (Charlesworth and Charlesworth, 1980; Kitano et al., 2009; but see Pennell et al., 2018; Anderson et al., 2020), which could provide an insight into sex chromosome evolution in taxa with old and highly differentiated sex chromosome systems such as insects or vertebrates (Blackmon et al. 2017; Stöck et al. 2021). It has been proposed that sex chromosome-autosome fusions are rare in vertebrate taxa with female heterogamety (Pokorná et al. 2014; Pennell et al. 2018). However, steadily growing number of neo-sex chromosomes have been reported in moth and butterflies (Lepidoptera), which comprise the most speciose group with female heterogamety (Nguyen and Carabajal Paladino 2016; Carabajal Paladino et al. 2019; Smith et al. 2019; Yoshido et al. 2020). Moreover, neo-sex chromosomes have been proposed to play a role in adaptive evolution and diversification of Lepidoptera (Nguyen et al. 2013; Smith et al. 2019; Yoshido et al. 2020). Sex chromosome-autosome fusions increase a number of sex-linked genes, which could significantly accelerate the accumulation of genetic incompatibilities between populations (Turelli and Begun 1997). Autosomes fused with sex chromosomes in large lepidopteran radiations were enriched for clusters of genes involved in detoxification and regulated absorption of plant secondary metabolites, which are crucial for larval performance on their host plants (Nguyen et al. 2013; Carabajal Paladino et al. 2019). Physical linkage between performance and sex-linked genes for either female host preference (Thompson 1988; Nygren et al. 2006) or host-independent reproductive isolation (Sperling 1994; Presgraves 2002) could facilitate adaptation and speciation in the presence of gene flow (Matsubayashi et al. 2010).

As recombination ceases due to achiasmatic meiosis in lepidopteran females, maternally transmitted (neo-W-linked) alleles of performance genes deteriorate which could contribute to population-specific divergence (Filatov 2018). Furthermore, selection for dosage compensation

caused by increasing environmental stress, is then expected and beneficial duplications of performance genes can be fixed by positive selection (Innan and Kondrashov 2010; Singh et al. 2020). Such amplification and functional divergence of performance genes upon sex chromosome-autosome fusions was proposed to be a key innovation, which enhanced adaptive radiation of leaf-rollers of superfamilies Tortricoidea and Gelechioidea (Nguyen et al. 2013; Carabajal Paladino et al. 2019).

Small ermine moths of the genus *Yponomeuta* (Yponomeutoidea, Yponomeutidae) have been a subject of multidisciplinary investigation of the evolution of insect-plant associations and speciation of phytophagous insects (Menken et al. 1992; Menken 1996; Menken and Roessingh 1998), and pheromone communication (Löfstedt et al. 1991; Liénard and Löfsted 2010). All ermine moths but one are monophagous and have an ancestral association with host plants of the family Celastraceae (Menken 1996). However, during their dispersion from East Asia to the western Palearctic, the common ancestor of the European clade switched its hosts from the family Celastraceae to Rosaceae and Salicaceae (Turner et al. 2010).

Analysis of meiotic nuclei of interspecific hybrids of the European ermine moths *Yponomeuta cagnagella* and *Y. padella* revealed numerous pairing irregularities including characteristic loops between paired homologues in meiosis, which indicate chromosome inversions (Hora et al. 2019). Saitoh (1960) reported male chromosome numbers of four *Yponomeuta* species. The karyotypes of *Y. malinella* and *Y. sedella* males corresponded to the ancestral lepidopteran karyotype n=31 exemplified by yponomeutoids *Atteva aurea* (Attevidae), *Zelleria haimbachi* (Yponomeutidae), and *Plutella xylostella* (Plutellidae) (Ennis, 1976; Kawazoé, 1987; cf. Van't Hof et al., 2013). By contrast, the chromosome number of *Y. polystictus* and *Y. sociatus* males was only n=30. The difference may be due to a sex chromosome-autosome fusion identified in females by Nilsson et al. (1988), who found a derived sex chromosome system Z_1Z_2W (2n♀=61, 2n♂=62) in six ermine moths of the *Y. cagnagellus–irrorellus* clade. The *Yponomeuta* ermine moths thus represent an ideal system for study a role of neo-sex chromosomes in ecological adaptation and speciation of Lepidoptera.

In the present study, we employed genomic *in situ* hybridization to identify sex chromosomes in 9 *Yponomeuta* spp. and their two outgroups, *Teinoptila gutella* (Yponomeutidae) and *Plutella xylostella* (Plutellidae). In *Y. evonymella*, we built genomic resources such as transcriptome sequence and a genomic library of bacterial artificial chromosomes (BACs) and used a custom designed microarray to identify Z-linked genes. Quantitative PCR (qPCR) with male and female genomic DNA (gDNA) and physical mapping by means of fluorescence *in situ* hybridization with BAC-derived probes confirmed that the *Y. evonymella* Z_2 chromosome corresponds to an autosome homeologous to the *Bombyx mori* chromosome 2 (BmChr2). The BmChr2 bears a cluster of genes with ovary-

specific expression which suggests that sexually antagonistic selection could have driven fixation of sex chromosome-autosome fusion in a common ancestor of *Yponomeuta* and *Teinoptila* spp.

Material and methods

Insects

All examined species were collected from wild populations except for *P. xylostella* which was obtain from a laboratory stock. The specimens were mostly collected as larvae and either processed immediately upon collection or reared under ambient conditions on their host plants (for details see **Suppl. Table S1**).

Chromosomal preparation

Meiotic and mitotic chromosomes were obtained from female and male gonads of 5th instar larvae by spreading technique as described in (Provazníková et al. 2021). Chromosomal preparations were afterwards dehydrated in ethanol series (70%, 80% and 100%, 30 sec each) and stored at -20°C or -80°C until further use.

RNA sequencing and transcriptome assembly

Total RNA was extracted from the *Y. evonymella* female larva with its gut removed using an RNA Blue reagent (Top-Bio, Prague, Czech Republic) following the manufacturer's protocol. The Illumina mRNA-seq library was constructed and sequenced on the Illumina HiSeq2000 platform by EMBL Genomics Core Facility (Heidelberg, Germany). Resulting raw 100-bp paired-end reads were trimmed and quality filtered by Trimmomatic version 0.30 ('LEADING:5 TRAILING:5 SLIDINGWINDOW:4:20'; Bolger et al., 2014). Transcriptome sequence was then assembled *de novo* by SOAPdenovo-trans-127mer (Xie et al. 2014) with multiple k-mer sizes ranging from 35 to 75 in increments of 10 and Trinity with the '--SS_lib_type RF' option (Haas et al. 2013). The resulting assemblies were merged and redundancy was removed using the EvidentialGene pipeline (Gilbert 2013). The raw reads were deposited in NCBI under SRA accession number PRJNA788289.

Array-CGH analysis

To identify sex-linked genes in *Y. evonymella*, we performed comparative genomic hybridization on a microarray (array-CGH) following (Baker and Wilkinson 2010). We searched for 1:1 orthologs of *B. mori* genes using the EvidentialGene dataset (see above) as input for HaMStR (Ebersberger et al. 2009) with the '-representative' option and lepidopteran core ortholog set by (Breinholt and Kawahara 2013). The *Y. evonymella* orthologs were used for design of 60-mer oligonucleotide

probes for a custom-made microarray slide using Agilent Technologies eArray design wizard (https://earray.chem.agilent.com/earray/). Female and male gDNA were extracted from larvae using CTAB protocol by (Winnepenninckx et al. 1993). DNA was quantified using the Qubit dsDNA BR Assay kit (Thermo Fisher Scientific, Waltham, MA, USA). DNA digestion, labelling, and array-CGH were performed by GenLabs (Prague, Czech Republic) according to a protocol for Agilent oligonucleotide array-based CGH for gDNA analysis. Hybridization intensities were extracted using Agilent's Feature Extraction software. Filtering and analysis of feature intensities followed (Baker and Wilkinson 2010) implemented in the custom Python script (Yoshido et al. 2020). The cut-off value of 0.5 was used to identify Z-linked genes.

***Y. evonymella* BAC library and screening**

Genomic library of BACs was constructed for *Y. evonymella* by AC Amplicon Express (Pullman, WA, USA). High molecular weight gDNA of *Y. evonymella* males was partially digested by *Hind* III and cloned into the pCC1BAC (Epicentre, Madison, WI, USA) vector, which was transformed into the DH10B *Escherichia coli* cells. The library consists of 20 736 clones with the average insert of about 125 Kbp. To identify BAC clones bearing genes of interest, i.e. a particular plate, row, and column of the BAC library, the library was screened by means of PCR according to the manufacturer's guidebook. For the screening procedure, individual BAC clones were pooled into 18 sets (or superpools) by the manufacturer, each containing 1 152 BAC clones. First, screening of superpools is performed to identify which contain BAC clone(s) with the sequence of interest. Each superpool corresponds to another subset of BACs comprising 21 matrixpools screened by second PCR. The matrixpools are plate, row, and column pools combined in a way, which allows identification of the clone bearing the gene of interest.

The 10µl PCR reaction contained 1x reaction buffer, 3 µM of each primer (**Suppl. Table S2),** 1 µl of template gDNA, 0.2 mM dNTPs and 0.5 U TaKaRa rTaq DNA polymerase (TaKaRa, Otsu, Japan). The amplification was carried out by PCR involving a denaturation step at 94 °C for 3 min; followed by 30 cycles of a denaturation at 94 °C for 30 s, an annealing at 58-60 °C for 45 s and an elongation at 72 °C for 45-60 s; and final elongation at 72 °C for 3 min. Afterwards, PCR reactions were visualized by gel electrophoresis and evaluated according to manufacturer's manual. To get single colonies, the positive BAC clones were plated on agar plate containing chloramphenicol (25 µg/ml). Presence of desired sequence was again verified by PCR on several single colonies per plate. PCR was prepared and carried out as described above using the single colonies as a template.

***In situ* hybridization**

Genomic DNA for fluorescence *in situ* hybridization (FISH) experiments was extracted from male and female larvae or pupae by standard phenol-chloroform method (Blin and Staford 1976). Obtained gDNA was amplified using illustra GenomiPhi HY DNA Amplification Kit (GE Healthcare, Milwaukee, WI, USA) according to manufacturer's protocol, purified by precipitation with isopropanol and sodium acetate, and dissolved in ultra clean water. DNA from selected BACs for was extracted by Plasmid Midi Kit (QIAGEN, Hilden, Germany) according to manufacturer's protocol.

Telomeric probe, female gDNA and BAC DNA were labelled by the nick translation using nick translation kit (Abbott Molecular Inc., Des Plaines, IL, USA) or a protocol described in (Kato et al. 2006) with some modifications (Dalíková et al. 2017). Using the nick translation kit (Abbott Molecular Inc.), the 25µl labelling reaction contained 500 ng DNA, 40 µM dATP, 40 µM dCTP, 40 µM dGTP, 14.4 µM dTTP and 25.6 µM Cy3-dUTP (Jena Bioscience, Jena, Germany) for telomeric probe; fluorescein-12-dUTP (Jena Biosciences) for female gDNA; Cy3-dUTP and fluorescein-12-dUTP (both Jena Biosciences) for BAC DNA. The reaction was incubated at 15°C for 75 min for telomeric probe, 4 h for female gDNA and 5 h for BAC DNA. The modified (Kato et al. 2006) nick translation reaction contained 1 µg of unlabelled DNA; 0.5 mM dATP, dCTP and dGTP; 0.1 mM dTTP; 20 µM of labelled nucleotides; 1x nick translation buffer (50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 0.005% BSA), 10 mM ß-mercaptoethanol, 2.5 x 10^{-4} U DNase I (ThermoFisher Scientific, Waltham, MA USA) and 1 U DNA polymerase I (ThermoFisher Scientific). The reaction was incubated at 15°C for 60 min to label telomeric probe, 210 min for female gDNA and 5 h for BAC DNA.

Genomic *in situ* hybridization (GISH) combined with telomeric probe and FISH with bacterial artificial chromosome (BAC-FISH) were carried out as described in (Yoshido et al., 2005a) and (Yoshido et al., 2005b), respectively, with some modifications. Briefly, the hybridization cocktail contained labelled probes, 500 ng fluorescein-labelled female gDNA and 100 ng of Cy3-labelled telomeric or 300 ng Cy3-labelled BAC DNA and 500 ng fluorescein-labelled BAC DNA, 3 µg of male competitor gDNA fragmented by heat for 20 min at 99°C, 25 µg sonicated salmon sperm DNA (Sigma-Aldrich, St. Louis, MO, USA) in 10 µl of 50% deionized formamide and 10% dextran sulfate in 2x SSC. The hybridization mixture was denatured for 5 min at 90°C. Chromosome slides were denatured in 70% formamide in 2x SSC for 3.5 min at 68°C. After 3-days hybridization at 37°C, slides were washed for 5 min in 0.1x SSC with 1% Triton X-100 at 62°C and counterstained with 0.5 µg/ml DAPI (4',6-diamidino-2-phenylindole; Sigma-Aldrich) in antifade with DABCO (1,4-diazabicyclo (2.2.2)-octane; Sigma-Aldrich).

Reprobing

To physically map multiple BACs, two rounds of BAC-FISH on the same chromosome slides were carried out due to the limiting number of available fluorochromes. The reprobing procedure was done as described in Zrzavá et al. (2018) with some modifications. The slides were incubated in 2x SSC for 30 min to remove cover slips and incubated for 10 min in 50% formamide, 1% Triton X in 0.1x SSC at 70°C to denature and eliminate the first probes. Afterwards, the slides were placed into prechilled 70% ethanol for 1 minute and then dehydrated at room temperature in 80% and 100% ethanol (30 s each). When dried, the slides were immediately used for another hybridization with BAC probes as described above.

Documentation and image processing

Preparations from FISH experiments were observed in a Zeiss Axioplan 2 microscope (Carl Zeiss, Jena, Germany) with a fluorescence filter sets and a monochrome CCD camera XM10 (Olympus Europa Holding, Hamburg, Germany). The images were captured in black-and-white, separately for each fluorescent dye with cellSens Standard software version 1.9 (Olympus). Subsequently, the images were pseudocolored and merged in Adobe Photoshop CS4 (version 11).

Quantitative PCR

To confirm results from array-CGH and BAC-FISH, and to verify a common origin of the Z_2 chromosome across the genus *Yponomeuta*, testing for sex-linkage of selected genes by quantitative PCR (qPCR) was performed in selected species, namely *Y. plumbella*, *Y. evonymella* and *Y. tokyonella.* Experiment was designed and carried out according to Nguyen et al. (2013) with some modifications. The qPCR analyses were performed with *Acetylcholinesterase 2 (Ace2)* as an autosomal reference, genes *Henna* and *Kettin* as markers for the ancestral Z_1 chromosome and genes *Arp6* and *Plep1* as markers for the Z_2 chromosome. The reference gene and genes of interest were analysed simultaneously in three biological and technical replicas for both males and females. For qPCR experiments, gDNA was extracted from male and female individuals of larval or pupal stage by NucleoSpin Tissue kit (Macherey-Nagel, Düren, Germany) or NucleoSpin Tissue XS kit (Macherey-Nagel) according to manufacturer's protocol. One 10µl reaction contained 1-10 ng of gDNA, 0.4 or 0.8 µM each primer (details in **Suppl. Tab. S3**) and 5 µL of SYBR Mix (Xceed qPCR SG 2x Mix Lo-ROX, IAB, Prague, Czech Republic). The experiment was carried out using the C1000 Thermal cycler CFX96 Real-Time System (Bio-Rad, Hercules, CA, USA). Data were analysed using software Bio-Rad CFX Manager 3.1. To determine the amplification efficiency of the reaction for each gene (*E*); 0x, 5x, 25x and 125x dilution series of gDNA pool from all gDNA samples of each species was analysed. Using the formula $R = [(1+ E_{\text{Reference}})^{\text{CtReference}}] / [(1+ E_{\text{Target}})^{\text{CtTarget}}]$ (Rovatsos et al. 2014), the target to reference gene dose ratio (*R*) was calculated for each biological sample. The statistical analysis was

carried out as described in Dalíková et al. (2017). Briefly, two hypothesis, autosomal hypothesis (*R* value ratio male:female to be 1:1) and Z-linkage hypothesis (*R* value ratio male:female to be 2:1) were tested by unpaired two-tailed t test for unequal variances.

Results

Karyotype analysis

Diploid chromosome numbers in *P. xylostella* (2n=62), *Y. cagnagella* (2n=♀61/♂62), *Y. padella* (2n=♀61/♂62) and *Y. evonymella* (2n=♀61/♂62) were already known from previous studies (Kawazoé 1987; Nilsson et al. 1988) and verified in this study. Based on mitotic metaphases and meiotic bivalents, a diploid chromosome number was determined to be 2n=♀61/♂62 in remaining species, namely *T. gutella*, *Y. plumbella*, *Y. polysticta*, *Y. kanaiella* and *Y. mahalebella* **(Fig. 1)** with exception of *Y. tokyonella* where reduced chromosome number was observed (2n=60) **(Fig. 2d)**. In *Y. orientalis*, a diploid chromosome number was also estimated to be 2n=♀61/♂62, however, due to lack of mitotic chromosome preparations, more nuclei and specimen need to be examined to verify the preliminary results **(Fig. 1i, j)**.

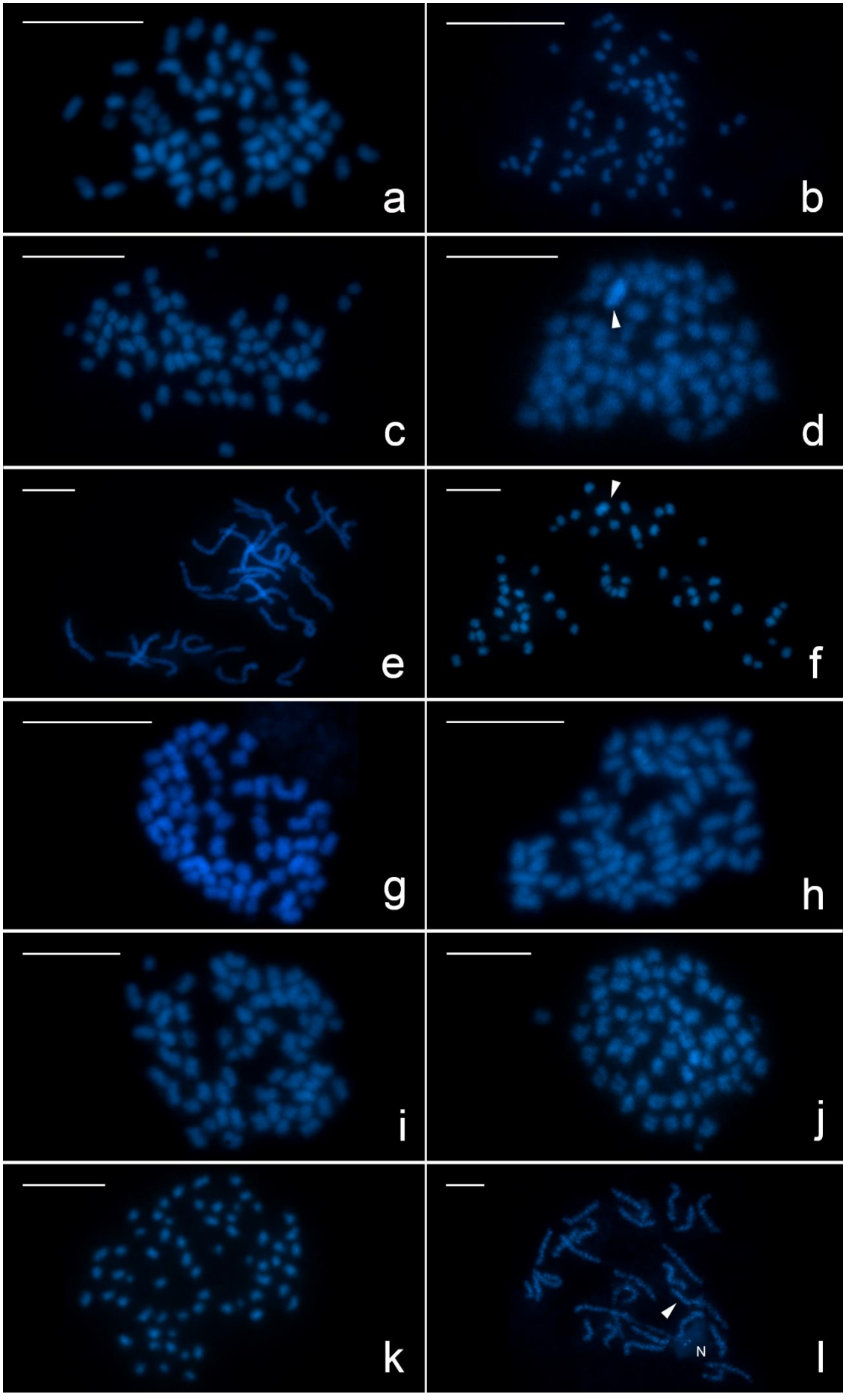
a
b
c
d
e
f
g
h
i
j
k
l
N

Figure 1: Male and female karyotypes of five *Yponomeuta* species and their outgroup *Teinoptila gutella.* Chromosomes are counterstained by DAPI (blue). a - male mitotic nuclei of *Teinoptila gutella* (2n=62), b - female mitotic nuclei of *Teinoptila gutella*(2n=61), c - male mitotic nuclei of *Yponomeuta plumbella* (2n=62), d - female mitotic nuclei of *Yponomeuta plumbella* (2n=61), e - male pachytene nuclei of *Y. polysticta* (2n=62), f - female mitotic nuclei of *Y. polysticta* (2n=61), g - male mitotic nuclei of *Y. kanaiella* (2n=62), h - female mitotic nuclei of *Y. kanaiella* (2n=61), i - male mitotic nuclei of *Y. orientalis* (2n=62), j - female mitotic nuclei of *Y. orientalis* (2n=61), k - male mitotic nuclei of *Y. mahalebella* (2n=62), k - female pachytene nuclei of *Y. mahalebella* (2n=61). N - nucleolus, W chromosome is indicated by arrowhead. Scale = 10 µm.

GISH with telomeric probe

To determine the constitution of sex chromosomes, genomic *in situ* hybridization (GISH) using female gDNA probe combined with telomeric probe was carried out in all studied species except for

Plutella xylostella where the sex chromosome system has been already known to be ZW/ZZ (Dalíková et al. 2017) The multiple sex chromosome constitution, ♀Z_1Z_2W/♂$Z_1Z_1Z_2Z_2$, was observed in all studied species except for *Y. tokyonella* with ♀ZW/♂ZZ system **(Fig. 2d**). In pachytene stage during meiosis in female ovaries both Z chromosomes, Z_1 and Z_2 chromosomes, pair with W chromosome forming a trivalent which is the biggest element of the karyotype. In *T. gutella, Y. plumbella, Y. kanaiella, Y. evonymella*, *Y. orientalis, Y. padella* and *Y. mahalebella*, the female gDNA probe hybridized evenly along the whole length of W chromosome (**Fig. 2a, b, e-g, i, j**). Whereas in *Y. cagnagella* and *Y. polysticta*, part of W chromosome was showed higher signal intensity which was probably caused by presence of compact heterochromatin of the ancestral W chromosome (**Fig. 2c, h**).

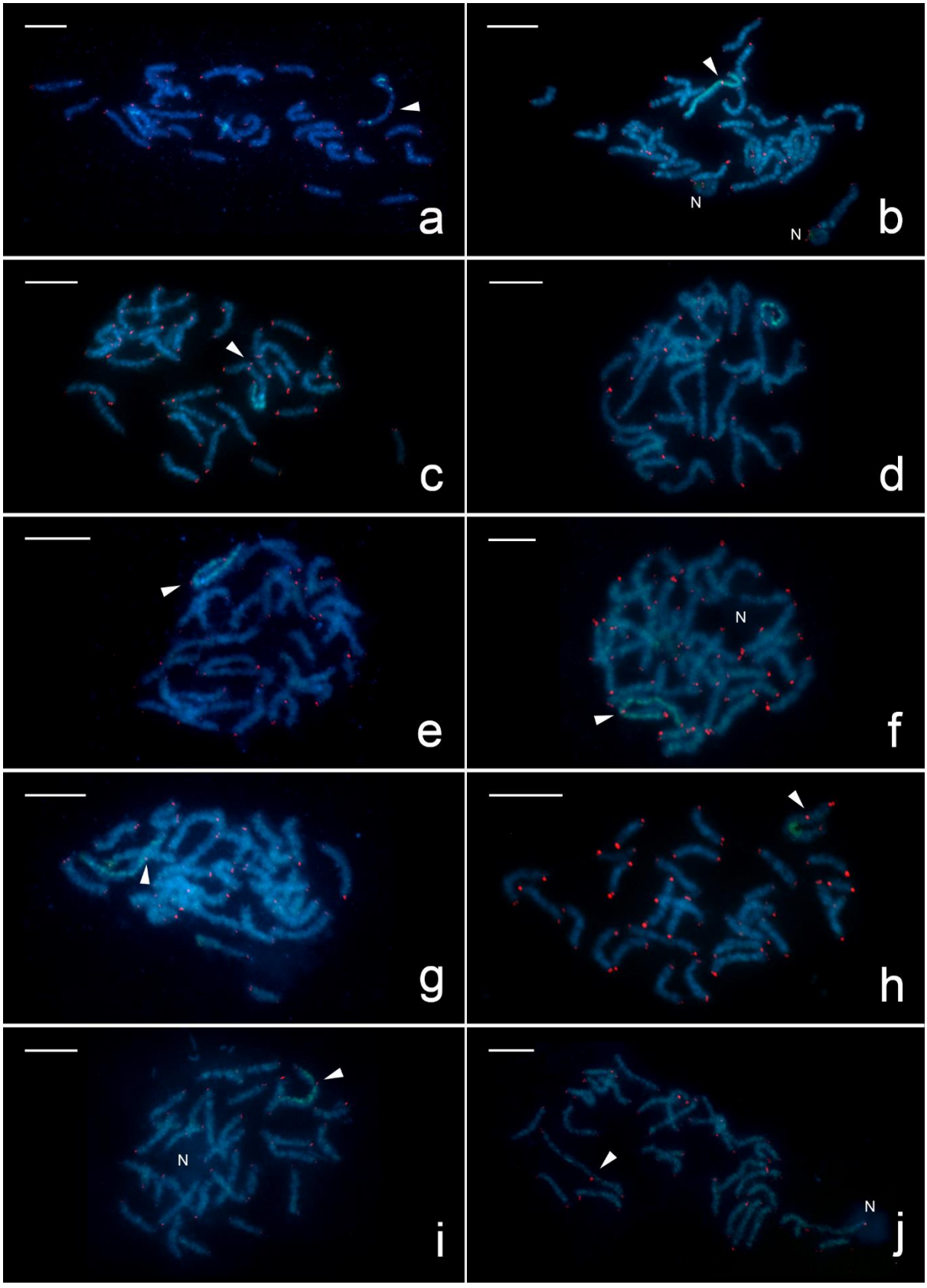

Figure 2: Sex chromosome trivalent (Z_1Z_2W) detected by GISH with female gDNA probe (green) and telomeric probe (red) on female pachytene nuclei (blue) of nine *Yponomeuta* spp. and their outgroup *Teinoptila gutella*. Telomeric signals of Z_1 and Z_2 chromosomes within the trivalent are marked by arrowheads. a - *Teinoptila gutella, b - Yponomeuta plumbella,* c - *Y. polysticta,* d - *Y.*

tokyonella, e- *Y. kanaiella,* f - *Y. evonymella,* g - *Y. orientalis,* h – *Y. cagnagella,* i - *Y. padella,* j - *Y. mahalebella.* N - nucleolus. Scale = 10 µm.

Array-CGH

To identify Z-linked orthologs in *Y. evonymella*, we carried out array-CGH (**Fig. 3**). After filtering, *Y. evonymella* $\log_2$ ratio values of male-to-female signal intensities ($\log_2$(M>F) for 4477 orthologs were obtained. The values averaged across two replicas clearly showed a bimodal distribution (**Fig. 3**). Using a cut-off value of 0.5, we identified 245 putative Z-linked orthologs. The orthologues were assigned to chromosomes assuming a conserved synteny of genes between *Y. evonymella* and *B. mori*. The identified *Y. evonymella* Z-linked orthologs were assigned to the *B. mori* Z chromosome (BmChr1) and chromosome 2 (BmChr2; **Fig. 3**).

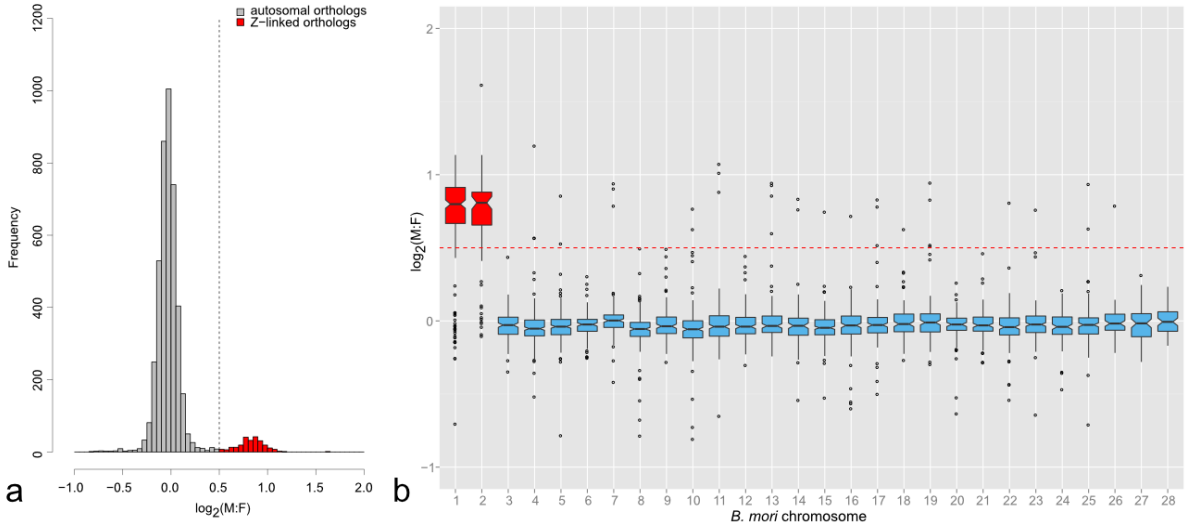

Figure 3: Array-CGH in *Y. evonymella*. a - Distribution of CGH $\log_2$ ratio of male-to-female signal intensities [$\log_2$(M:F)]. The peak centered at −0.05 (grey) corresponds to putative autosomal orthologs while bins with values >0.5 (dashed line) form the smaller peak comprise putative Z-linked orthologs (red). A total of 4477 orthologs are presented. b – Assignment of *Y. evonymella* orthologs to the *B. mori* chromosomes. The putative *Y. evonymella* Z-linked orthologues were assigned to chromosomes 1 (Z) and 2 of *B. mori* (red) with $\log_2$(M:F)>0.5 (dashed line). Autosomes are in blue. Boxes represent median and first and third quartiles. Whiskers extend to 1.5 * IQR from the hinges, where IQR is distance between the first and third quartiles.

BAC-FISH

To verify results of the array-CGH analysis, we physically mapped chromosome markers to the *Y. evonymella* Z_1 and Z_2 chromosomes. As putative markers for the chromosome Z_1, the *Y. evonymella* BAC library was screened for BACs containing single copy orthologs to genes *Henna* and *Kettin*, in *B. mori* localized on chromosome Z (BmChr1). Analogously, BACs containing orthologs of *Arp6* and Plep1 linked to chromosome 2 in *B. mori* (BmChr2) were used for the chromosome Z_2. The BAC clones bearing these markers were hybridized to female meiotic nuclei of *Y. evonymella* using BAC-

FISH (**Fig. 4**). BAC probes bearing Z_1 marker *Henna* and Z_2 markers *Arp6* and *Plep1* successfully hybridized to the sex chromosome trivalent and provided clear and discreet signals on the Z_1 and Z_2 chromosomes, respectively. The probe derived from BAC bearing *Kettin* hybridized to the terminal regions of all chromosomes (results not shown) which could be explained by presence of telomeric or other repetitive sequences in the BAC. The *Kettin* clone was therefore excluded from this analysis. This experiment corroborated results of array-CGH analysis and confirmed that the Z_1 and Z_2 chromosomes form together with the W chromosome a trivalent in *Y. evonymella*.

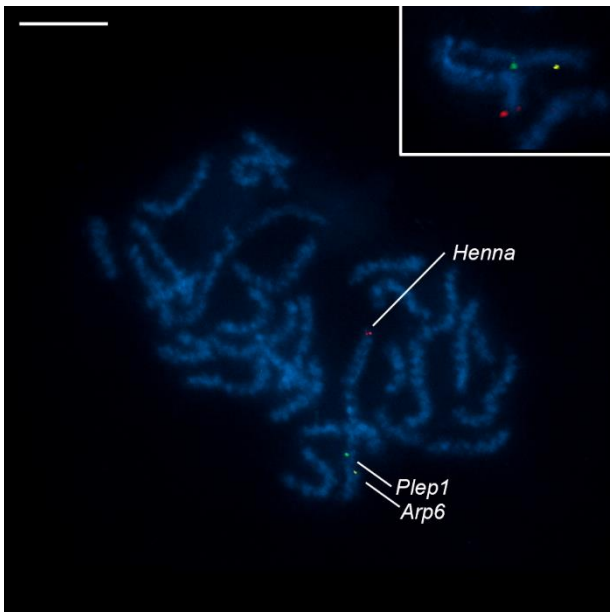

Figure 4: Localization of BAC clone bearing gene *Henna*, marker for Z_1 chromosome (red signal), and BAC clones bearing gene *Plep 1* (green signal) and *Arp 6* (yellow signal), markers for Z_2 chromosome detected on female pachytene nuclei of *Y. evonymella*. Chromosomes are counterstained by DAPI (blue). Hybridization signals are indicated by arrowhead. Scale = 10 µm.

Quantitative PCR

To verify the hypothesis that the multiple sex chromosome system occurred in the common ancestor of the genus *Yponomeuta,* a relative gene dose of the Z_1-linked genes, *Henna* and *Kettin,* and the Z_2-linked genes, *Arp6* and *Plep1*, was compared by qPCR experiment between female and male gDNA of *Y. evonymella,* and *Y. plumbella* representing the early diverged *Yponomeuta* species. We also tested *Y. tokyonella* in which chromosome number was reduced to 2n=60 in both sexes. The results showed statistically significant twofold difference between males and females in all studied genes and thus proved their Z-linkage in all three *Yponomeuta* species (**Fig. 5, Suppl. Tab. S4**). The only exception was the gene *Henna* in *Y. plumbella* which showed autosomal linkage. The Z chromosome is considered conserved across Lepidoptera and the Z-linkage of the *Henna* gene has been confirmed in many species (Van't Hof et al. 2013; Nguyen et al. 2013; Dalíková et al. 2017). It is

reasonable to assume that a chromosomal rearrangement such as translocation moved the *Henna* gene to an autosome.

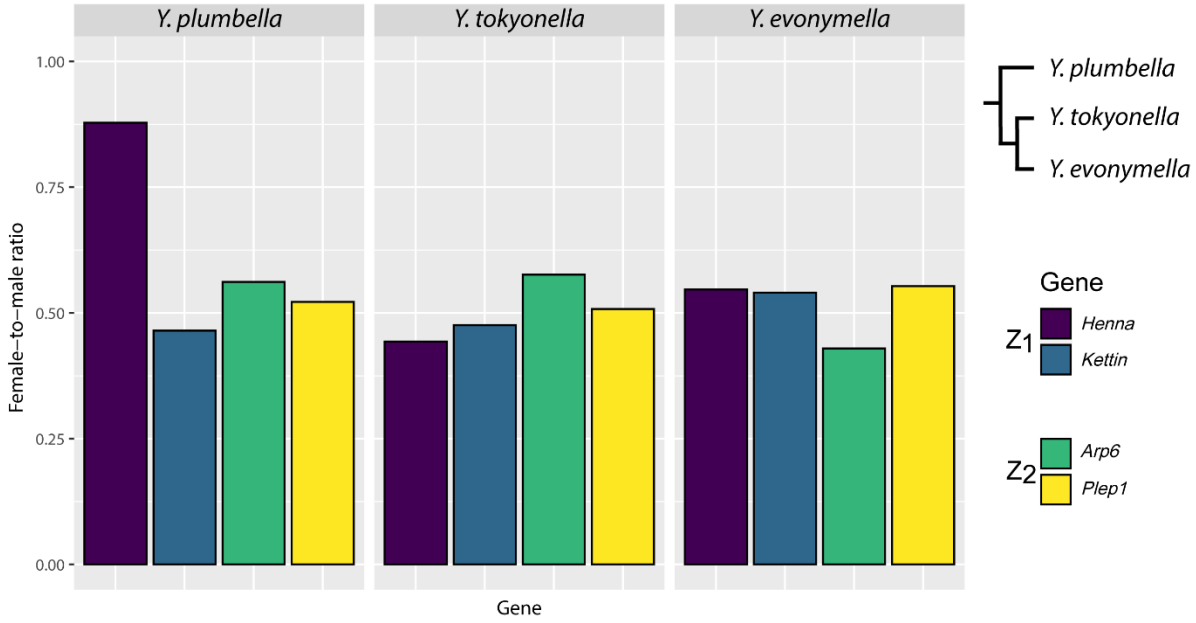

Figure 5: Mean female-to-male ratio values obtained by qPCR for Z_1-linked genes *Henna* and *Kettin* and Z_2-linked genes *Arp6* and *Plep1* using *Ace2* as the reference gene in *Y. plumbella*, *Y. tokyonella* and *Y. evonymella*. A value of 0.5 is expected for Z-linked genes, while for autosomal genes the expected value is 1. For summary of qPCR results, see **Supplementary Table S4**.

Discussion

Ermine moths of the genus *Yponomeuta* represent a suitable system to study a role of changes in genome organization in a host shift. While *Yponomeuta* and *Teinoptila* spp. have an ancestral association with a single plant family, Celastraceae, the European *Yponomeuta* clade shifted to new hosts of the families Rosaceae and Salicaceae. Using the combination of GISH and FISH with telomeric probe, we detected the heterologous W chromosome in all species under study and clearly identified the multiple sex chromosome constitution ♀Z_1Z_2W/♂$Z_1Z_1Z_2Z_2$ reported by (Nilsson et al. 1988) except for *Y. tokyonella* (**Fig. 2**). Our results thus show that formation of the multiple sex chromosomes probably not played a major role in the host shift of European ermine moths as they predate the split of the *Yponomeuta* and *Teinoptila* genera. Since a representative of a sister clade of ermine moths, *P. xylostella* (Plutellidae), has an ancestral genome organization 2n=62 with the ZW sex chromosome system (Ward et al. 2021), the *Yponomeuta* multiple sex chromosomes most likely rose within the family Yponomeutidae. Analysis of additional yponomeutids is necessary to pinpoint their exact origin. Results of array-CGH confirmed by physical mapping and qPCR further revealed that the Z_2 chromosome of the *Yponomeuta* multiple sex chromosome system corresponds to a synteny block homeologous to the *B. mori* chromosome 2 (BmChr2; **Fig. 3, 4, 5**).

In *Y. tokyonella*, we observed the ♀ZW/♂ZZ sex chromosome constitution, where the fusion between ancestral sex chromosomes and a pair of autosomes was completed, forming neo-W and neo-Z chromosome (♀WZ/♂ZZ; **Fig. 2d**). It was shown that autosomal fusions are not random in Lepidoptera, with small and repeat rich autosomes being repeatedly involved in distant taxa (Ahola et al. 2014). Furthermore, it was proposed that sex chromosome-autosome fusions are the first rearrangements to differentiate karyotypes from ancestral lepidopteran genome architecture (Carabajal Paladino et al. 2019). We hypothesize that sex chromosome turnover is initiated by the repeat rich W chromosome in Lepidoptera giving rise to the Z_1Z_2W multiple sex chromosome system such as the one observed in *Yponomeuta*. Pairing of the Z_1 and Z_2 chromosomes with the W chromosome increases probability of their interaction (cf. Schlecht et al. 2004), which can result in a fusion mediated by ectopic recombination and produce a conspicuously large pair of neo-sex chromosome observed in other Lepidoptera (Nguyen et al. 2013; Mongue et al. 2017; Carabajal Paladino et al. 2019).

According to the null model of sex chromosome-autosome fusions developed by Anderson et al. (2020), all chromosomes fuse with equal probability. The model showed that sex chromosome-autosome fusions make a large proportion of fusions even in absence of selection in species with small number of autosomes, while they should be rare in clades with high chromosome number such as Lepidoptera. Yet, growing number of neo-sex chromosome systems have been reported in Lepidoptera, which suggests that sex chromosome-autosome fusions are common in this female heterogametic group (Nguyen and Carabajal Paladino 2016; Carabajal Paladino et al. 2019). This is in stark contrast to analyses performed in vertebrates (Pokorná et al. 2014; Pennell et al. 2015; Sember et al. 2021), where fusions between sex chromosomes and autosomes are rare.

It was hypothesized that sex chromosome-autosome fusions are more likely to be deleterious compared to fusions between autosomes in species with achiasmatic meiosis as a result of sex chromosome differentiation process (Anderson et al. 2020). As genes cease to recombine upon a sex chromosome-autosome fusion, they start accumulating mutations. Deleterious mutations thus quickly overcome any initial fitness benefit of the fusion and prevent its fixation (Anderson et al., 2020; cf. Lenormand and Roze, 2022). The deleterious effect should be proportionate to number of genes born by the involved autosome. Thus, the high incidence of neo-sex chromosomes in Lepidoptera could be explained by their achiasmatic female meiosis and numerous small chromosomes.

The *B. mori* chromosome 2 is one of the smallest elements in the karyotype (Yoshido et al. 2005a). As mentioned above, smaller chromosomes with high repetitive DNA content are more

prone to chromosome rearrangements compared to large ones which generally have less repetitive DNA (Ahola et al. 2014). A recent analysis of repetitive landscape in two *Danaus* species showed that various mobile elements are more abundant in small chromosomes (Baril and Hayward 2022). Indeed, the chromosome 2 fused with other chromosomes in *S. cynthia* (2n = 25-28) (Yoshido et al. 2011) or *Manduca sexta* (n=28) (Yasukochi et al. 2009). Notably, a similar pattern was also observed in birds. The typical avian karyotype consists of about 80 macro- and micro- chromosomes (Ellegren 2010; Zhang et al. 2014) except for some groups such as parrots (de Oliveira Furo et al. 2017) and birds of prey (de Oliveira et al. 2005; Joseph et al. 2018). Their reduced chromosome numbers are mostly caused by lineage specific fusions of microchromosomes and smaller macrochromosomes including sex chromosomes (Wilcox et al. 2019; Furo et al. 2020; Huang et al. 2022). It was hypothesized that in parrots, expansion of mobile elements could lead to chromosome rearrangements and subsequent loss of genes involved in maintaining genome stability and repair of double-strand breaks (Huang et al. 2022).

Alternatively, it was proposed that fixation of neo-sex chromosomes in lepidopteran populations could be facilitated by sexual antagonistic selection (Charlesworth and Charlesworth 1980; Smith et al. 2016; Matsumoto and Kitano 2016) or selection for linkage between largely sex-linked reproductive barriers and larval performance (Nguyen et al. 2013; Carabajal Paladino et al. 2019). As for the latter, the inspection of a gene content of the chromosome 2 in the reference genome of *B. mori* did not show any enrichment for genes involved in detoxification of plant secondary metabolites, which are crucial for performance of larvae on their host plants (cf. Yu et al., 2008; Tsubota and Shiotsuki, 2010; Ai et al., 2011; Ahn et al., 2012; Xie et al., 2012).

However, BmChr2 bears a cluster of more than 100 chorion genes, which are expressed in ovaries and encode specialized structural proteins found in the eggshell (Goldsmith and Basehoar 1978; Suetsugu et al. 2013). Notably, (Suetsugu et al. 2013) shown that 74% of genes with ovary specific expression is clustered on chromosomes 2, 10, 15, and 16 in the *B. mori* genome. While a fragment of chromosome 2 fused with a Z chromosome also in *Pieris* white butterflies (Pieridae; Hill et al., 2019; Steward et al., 2021), chromosomes 15 and 16 fused with Z chromosomes in moths of the family Tortricidae and *Danaus* spp., respectively (Nguyen et al. 2013; Mongue et al. 2017). Ovary specific expression and sex linkage are alternative ways to resolve sexual conflict (Mank 2009). Thus, we hypothesize that sex chromosome turnover in Lepidoptera could be driven by sexual antagonism, which is in agreement with theoretical predictions (Charlesworth and Charlesworth 1980; Matsumoto and Kitano 2016) and findings in a three-spined stickleback fish (Kitano et al. 2009; Dagilis et al. 2022). Further research on distribution of sexual antagonistic loci in lepidopteran genomes is needed to test this hypothesis.

Literature:

Ahn S-J, Vogel H, Heckel DG (2012) Comparative analysis of the UDP-glycosyltransferase multigene family in insects. Insect Biochem Mol Biol 42:133–147. https://doi.org/10.1016/j.ibmb.2011.11.006

Ahola V, Lehtonen R, Somervuo P, et al (2014) The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nat Commun 5:4737. https://doi.org/10.1038/ncomms5737

Ai Q, Xu H, Mai K, et al (2011) Effects of dietary supplementation of *Bacillus subtilis* and *fructooligosaccharide* on growth performance, survival, non-specific immune response and disease resistance of juvenile large yellow croaker, *Larimichthys crocea*. Aquaculture 317:155–161. https://doi.org/10.1016/j.aquaculture.2011.04.036

Anderson NW, Hjelmen CE, Blackmon H (2020) The probability of fusions joining sex chromosomes and autosomes. Biol Lett 16:20200648. https://doi.org/10.1098/rsbl.2020.0648

Baker RH, Wilkinson GS (2010) Comparative genomic hybridization (CGH) reveals a neo-X chromosome and biased gene movement in Stalk-eyed flies (Genus Teleopsis). PLoS Genet 6:e1001121. https://doi.org/10.1371/journal.pgen.1001121

Baril T, Hayward A (2022) Migrators within migrators: exploring transposable element dynamics in the monarch butterfly, *Danaus plexippus*. Mobile DNA 13:5. https://doi.org/10.1186/s13100-022-00263-5

Blackmon H, Ross L, Bachtrog D (2017) Sex determination, sex chromosomes, and karyotype evolution in insects. J Heredity 108:78–93. https://doi.org/10.1093/jhered/esw047

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Breinholt JW, Kawahara AY (2013) Phylotranscriptomics: saturated third codon positions radically influence the estimation of trees based on next-gen data. Genome Biol Evol 5:2082–2092. https://doi.org/10.1093/gbe/evt157

Carabajal Paladino LZ, Provazníková I, Berger M, et al (2019) Sex chromosome turnover in moths of the diverse superfamily Gelechioidea. Genome Biol Evol 11:1307–1319. https://doi.org/10.1093/gbe/evz075

Charlesworth D, Charlesworth B (1980) Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genet Res 35:205–214. https://doi.org/10.1017/S0016672300014051

Connallon T, Olito C, Dutoit L, et al (2018) Local adaptation and the evolution of inversions on sex chromosomes and autosomes. Phil Trans R Soc B 373:20170423. https://doi.org/10.1098/rstb.2017.0423

Dagilis AJ, Sardell JM, Josephson MP, et al (2022) Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes. Phil Trans R Soc B 377:20210205. https://doi.org/10.1098/rstb.2021.0205

Dalíková M, Zrzavá M, Hladová I, et al (2017) New insights into the evolution of the W chromosome in Lepidoptera. J Hered 108:709–719. https://doi.org/10.1093/jhered/esx063

de Oliveira EHC, Habermann FA, Lacerda O, et al (2005) Chromosome reshuffling in birds of prey: the karyotype of the world's largest eagle (Harpy eagle, *Harpia harpyja*) compared to that of the chicken (*Gallus gallus*). Chromosoma 114:338–343. https://doi.org/10.1007/s00412-005-0009-5

de Oliveira Furo I, Kretschmer R, dos Santos MS, et al (2017) Chromosomal mapping of repetitive DNAs in *Myiopsitta monachus* and *Amazona aestiva* (Psittaciformes, Psittacidae) with emphasis on the sex chromosomes. Cytogenet Genome Res 151:151–160. https://doi.org/10.1159/000464458

Ebersberger I, Strauss S, von Haeseler A (2009) HaMStR: Profile hidden markov model based search for orthologs in ESTs. BMC Evol Biol 9:157. https://doi.org/10.1186/1471-2148-9-157

Ellegren H (2010) Evolutionary stasis: the stable chromosomes of birds. Trends Ecol Evol 25:283–291. https://doi.org/10.1016/j.tree.2009.12.004

Ennis TJ (1976) Sex chromatin and chromosomes numbers in Lepidoptera. Can J Genet Cytol 18:119–130. https://doi.org/10.1139/g76-017

Filatov DA (2018) The two "rules of speciation" in species with young sex chromosomes. Mol Ecol 27:3799–3810. https://doi.org/10.1111/mec.14721

Furo I de O, Kretschmer R, O'Brien PC, et al (2020) Chromosomal Evolution in the Phylogenetic Context: A Remarkable Karyotype Reorganization in Neotropical Parrot *Myiopsitta monachus* (Psittacidae). Front Genet 11:721. https://doi.org/10.3389/fgene.2020.00721

Gilbert D (2013) Gene-omes built from mRNA seq not genome DNA. 7th Annual Arthropod Genomics Symposium, Notre Dame, France. https://doi.org/10.7490/f1000research. 1112594.1

Goldsmith MR, Basehoar G (1978) Organization of the chorion genes of *Bombyx mori*, a multigene family. I. evidence for linkage to chromosome *2*. Genetics 90:291–310. https://doi.org/10.1093/genetics/90.2.291

Haas BJ, Papanicolaou A, Yassour M, et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. https://doi.org/10.1038/nprot.2013.084

Hill J, Rastas P, Hornett EA, et al (2019) Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. Sci Adv 5:eaau3648. https://doi.org/10.1126/sciadv.aau3648

Hora KH, Marec F, Roessingh P, Menken SBJ (2019) Limited intrinsic postzygotic reproductive isolation despite chromosomal rearrangements between closely related sympatric species of small ermine moths (Lepidoptera: Yponomeutidae). Biol J Linn Soc 128:44–58. https://doi.org/10.1093/biolinnean/blz090

Huang Z, De O. Furo I, Liu J, et al (2022) Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. Nat Commun 13:944. https://doi.org/10.1038/s41467-022-28585-1

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11:97–108. https://doi.org/10.1038/nrg2689

Joseph S, O'Connor R, Al Mutery A, et al (2018) Chromosome level genome assembly and comparative genomics between three falcon species reveals an unusual pattern of genome organisation. Diversity 10:113. https://doi.org/10.3390/d10040113

Kato A, Kato A, Albert P, et al (2006) Sensitive fluorescence *in situ* hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. Biotechnic & Histochemistry 81:71–78. https://doi.org/10.1080/10520290600643677

Kawazoé A (1987) The chromosome in the primitive or microlepidopterous moth-groups II. Proc Jpn Acad, Ser B 63:87–90. https://doi.org/10.2183/pjab.63.87

Kitano J, Ross JA, Mori S, et al (2009) A role for a neo-sex chromosome in stickleback speciation. Nature 461:1079–1083. https://doi.org/10.1038/nature08441

Kratochvíl L, Gamble T, Rovatsos M (2021) Sex chromosome evolution among amniotes: is the origin of sex chromosomes non-random? Phil Trans R Soc B 376:20200108. https://doi.org/10.1098/rstb.2020.0108

Lenormand T, Roze D (2022) Y recombination arrest and degeneration in the absence of sexual dimorphism. Science 375:663–666. https://doi.org/10.1126/science.abj1813

Liénard MA, Löfsted C (2010) Functional flexibility as a prelude to signal diversity? Role of a fatty acyl reductase in moth pheromone evolution: R\role of a fatty acyl reductase in moth pheromone evolution. Comm Integ Biol 3:586–588. https://doi.org/10.4161/cib.3.6.13177

Löfstedt C, Herrebout WM, Menken SBJ (1991) Sex pheromones and their potential role in the evolution of reproductive isolation in small ermine moths (Yponomeutidae). Chemoecology 2:20–28. https://doi.org/10.1007/BF01240662

Mank JE (2009) The W, X, Y and Z of sex-chromosome dosage compensation. Trends in Genetics 25:226–233. https://doi.org/10.1016/j.tig.2009.03.005

Matsubayashi KW, Ohshima I, Nosil P (2010) Ecological speciation in phytophagous insects. Entomol Exp Appl 134:1–27. https://doi.org/10.1111/j.1570-7458.2009.00916.x

Matsumoto T, Kitano J (2016) The intricate relationship between sexually antagonistic selection and the evolution of sex chromosome fusions. J Theor Biol 404:97–108. https://doi.org/10.1016/j.jtbi.2016.05.036

Menken SBJ (1996) Pattern and process in the evolution of insect-plant associations: *Yponomeuta* as an example. Entomol Exp Appl 80:297–305. https://doi.org/10.1111/j.1570-7458.1996.tb00940.x

Menken SBJ, Herrebout WM, Wiebes JT (1992) Small ermine moths (Yponomeuta): their host relations and evolution. Annu Rev Entomol 37:41–66. https://doi.org/10.1146/annurev.en.37.010192.000353

Menken SBJ, Roessingh P (1998) Endless Forms: Species and Speciation. In: Evolution of insect-plant associations: sensoty perception and receptor modifications direct food specialization and host shifts in phytophagous insects. Academic Press, New York, pp 145–156

Mongue AJ, Nguyen P, Voleníková A, Walters JR (2017) Neo-sex chromosomes in the Monarch butterfly, *Danaus plexippus*. 3G-Genes Genom Genet 7:3281–3294. https://doi.org/10.1534/g3.117.300187

Nguyen P, Carabajal Paladino LZ (2016) Evolutionary biology: convergent evolution, evolution of complex traits, concepts and methods. In: On the neo-sex chromosomes of Lepidoptera. Springer, Heidelberg, pp 171–185

Nguyen P, Sýkorová M, Šíchová J, et al (2013) Neo-sex chromosomes and adaptive potential in tortricid pests. PNAS 110:6931–6936. https://doi.org/10.1073/pnas.1220372110

Nilsson N-O, 'Löfstedt C, Dävring L (1988) Unusual sex chromosome inheritance in six species of small ermine moths (Yponomeuta, Yponomeutidae, Lepidoptera). Hereditas 108:259–265. https://doi.org/10.1111/j.1601-5223.1988.tb00311.x

Nygren GH, Nylin S, Stefanescu C (2006) Genetics of host plant use and life history in the comma butterfly across Europe: varying modes of inheritance as a potential reproductive barrier. J Evolution Biol 19:1882–1893. https://doi.org/10.1111/j.1420-9101.2006.01174.x

Payseur BA, Presgraves DC, Filatov DA (2018) Introduction: Sex chromosomes and speciation. Mol Ecol 27:3745–3748. https://doi.org/10.1111/mec.14828

Pennell MW, Kirkpatrick M, Otto SP, et al (2015) Y Fuse? Sex chromosome fusions in fishes and reptiles. PLoS Genet 11:e1005237. https://doi.org/10.1371/journal.pgen.1005237

Pennell MW, Mank JE, Peichel CL (2018) Transitions in sex determination and sex chromosomes across vertebrate species. Mol Ecol 27:3950–3963. https://doi.org/10.1111/mec.14540

Pokorná M, Altmanová M, Kratochvíl L (2014) Multiple sex chromosomes in the light of female meiotic drive in amniote vertebrates. Chromosome Res 22:35–44. https://doi.org/10.1007/s10577-014-9403-2

Presgraves DC (2002) Patterns of postzygotic isolation in Lepidoptera. Evolution 56:1168–1183. https://doi.org/10.1111/j.0014-3820.2002.tb01430.x

Provazníková I, Hejníčková M, Visser S, et al (2021) Large-scale comparative analysis of cytogenetic markers across Lepidoptera. Sci Rep 11:12214. https://doi.org/10.1038/s41598-021-91665-7

Rovatsos M, Altmanová M, Pokorná M, Kratochvíl L (2014) Conserved sex chromosomes across adaptively radiated *Anolis lizards*. Evolution 68:2079–2085. https://doi.org/10.1111/evo.12357

Saitoh K (1960) A chromosome survey in thirty species of moths. Jap J Genet 35:41–48

Schlecht HB, Lichten M, Goldman ASH (2004) Compartmentalization of the yeast meiotic nucleus revealed by analysis of ectopic recombination. Genetics 168:1189–1203. https://doi.org/10.1534/genetics.104.029157

Sember A, Nguyen P, Perez MF, et al (2021) Multiple sex chromosomes in teleost fishes from a cytogenetic perspective: state of the art and future challenges. Phil Trans R Soc B 376:20200098. https://doi.org/10.1098/rstb.2020.0098

Singh KS, Troczka BJ, Duarte A, et al (2020) The genetic architecture of a host shift: an adaptive walk protected an aphid and its endosymbiont from plant chemical defenses. Sci Adv 6:eaba1070. https://doi.org/10.1126/sciadv.aba1070

Smith DAS, Gordon IJ, Traut W, et al (2016) A neo-W chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. Proc R Soc B 283:20160821. https://doi.org/10.1098/rspb.2016.0821

Smith DAS, Traut W, Martin SH, et al (2019) Neo sex chromosomes, colour polymorphism and male-killing in the African Queen Butterfly, *Danaus chrysippus* (L.). Insects 10:291. https://doi.org/10.3390/insects10090291

Sperling FAH (1994) Sex-linked genes and species differences in Lepidoptera. Can Entomol 126:807–818. https://doi.org/10.4039/Ent126807-3

Steward RA, Okamura Y, Boggs CL, et al (2021) The genome of the margined white butterfly (*Pieris macdunnoughii*): sex chromosome insights and the power of polishing with PoolSeq data. Genome Biol Evol 13:evab053. https://doi.org/10.1093/gbe/evab053

Stöck M, Kratochvíl L, Kuhl H, et al (2021) A brief review of vertebrate sex evolution with a pledge for integrative research: towards ‘ *sexomics* .’ Phil Trans R Soc B 376:20200426. https://doi.org/10.1098/rstb.2020.0426

Suetsugu Y, Futahashi R, Kanamori H, et al (2013) Large scale full-length cDNA sequencing reveals a unique genomic landscape in a lepidopteran model insect, *Bombyx mori*. G3-Genes Genom Genet 3:1481–1492. https://doi.org/10.1534/g3.113.006239

Thompson JN (1988) Evolutionary genetics of oviposition preference in swallowtail butterflies. Evolution 42:1223–1234. https://doi.org/10.1111/j.1558-5646.1988.tb04182.x

Tsubota T, Shiotsuki T (2010) Genomic analysis of carboxyl/cholinesterase genes in the silkworm *Bombyx mori*. BMC Genomics 11:377. https://doi.org/10.1186/1471-2164-11-377

Turelli M, Begun DJ (1997) Haldane’s Rule and X-chromosome Size in Drosophila. Genetics 147:1799–1815. https://doi.org/10.1093/genetics/147.4.1799

Turner H, Lieshout N, Van Ginkel WE, Menken SBJ (2010) Molecular phylogeny of the small ermine moth genus *Yponomeuta* (Lepidoptera, Yponomeutidae) in the Palaearctic. PLoS ONE 5:e9933. https://doi.org/10.1371/journal.pone.0009933

Van’t Hof AE, Nguyen P, Dalíková M, et al (2013) Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a model of industrial melanism. Heredity 110:283–295. https://doi.org/10.1038/hdy.2012.84

Ward CM, Perry KD, Baker G, et al (2021) A haploid diamondback moth (*Plutella xylostella* L.) genome assembly resolves 31 chromosomes and identifies a diamide resistance mutation. Insect Biochem Mol Biol 138:103622. https://doi.org/10.1016/j.ibmb.2021.103622

Wilcox JJS, Boissinot S, Idaghdour Y (2019) Falcon genomics in the context of conservation, speciation, and human culture. Ecol Evol 9:14523–14537. https://doi.org/10.1002/ece3.5864

Wright AE, Dean R, Zimmer F, Mank JE (2016) How to make a sex chromosome. Nat Commun 7:12087. https://doi.org/10.1038/ncomms12087

Xie X, Cheng T, Wang G, et al (2012) Genome-wide analysis of the ATP-binding cassette (ABC) transporter gene family in the silkworm, *Bombyx mori*. Mol Biol Rep 39:7281–7291. https://doi.org/10.1007/s11033-012-1558-3

Xie Y, Wu G, Tang J, et al (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666. https://doi.org/10.1093/bioinformatics/btu077

Yasukochi Y, Tanaka-Okuyama M, Shibata F, et al (2009) Extensive conserved synteny of genes between the karyotypes of *Manduca sexta* and *Bombyx mori* revealed by BAC-FISH mapping. PLoS ONE 4:e7465. https://doi.org/10.1371/journal.pone.0007465

Yoshido A, Bando H, Yasukochi Y, Sahara K (2005a) The *Bombyx mori* karyotype and the assignment of linkage groups. Genetics 170:675–685. https://doi.org/10.1534/genetics.104.040352

Yoshido A, Marec F, Sahara K (2005b) Resolution of sex chromosome constitution by genomic in situ hybridization and fluorescence in situ hybridization with (TTAGG)n telomeric probe in some species of Lepidoptera. Chromosoma 114:193–202

Yoshido A, Šíchová J, Pospíšilová K, et al (2020) Evolution of multiple sex-chromosomes associated with dynamic genome reshuffling in Leptidea wood-white butterflies. Heredity 125:138–154. https://doi.org/10.1038/s41437-020-0325-9

Yoshido A, Yasukochi Y, Sahara K (2011) *Samia cynthia* versus *Bombyx mori*: comparative gene mapping between a species with a low-number karyotype and the model species of Lepidoptera. Insect Biochem Mol Biol 41:370–377. https://doi.org/10.1016/j.ibmb.2011.02.005

Yu Q, Lu C, Li B, et al (2008) Identification, genomic organization and expression pattern of glutathione S-transferase in the silkworm, *Bombyx mori*. Insect Biochem Mol Biol 38:1158–1164. https://doi.org/10.1016/j.ibmb.2008.08.002

Zhang G, Li C, Li Q, et al (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311–1320. https://doi.org/10.1126/science.1251385

Zrzavá M, Hladová I, Dalíková M, et al (2018) Sex chromosomes of the iconic moth *Abraxas grossulariata* (Lepidoptera, Geometridae) and its congener *A. sylvata*. Genes 9:279. https://doi.org/10.3390/genes9060279

Supplementary tables

Supplementary table S1: List of species examined.

Species, populations	Family	Origin	Condition, diet
Plutella xylostella	Plutellidae	Laboratory strain*	25 ± 1°C, 16/8 h (light/dark) regime, artificial diet [1]
Teinoptila gutella	Yponomeutidae	Yaese, Okinawa, Japan	-
Yponomeuta plumbella	Yponomeutidae	Meijendel, Netherlands	*Euonymus europaeus*
Yponomeuta kanaiella	Yponomeutidae	Bibi, Hokkaido, Japan	-
Yponomeuta polysticta	Yponomeutidae	Iwamizawa, Hokkaido, Japan	-
Yponomeuta tokyonella	Yponomeutidae	Iwamizawa, Hokkaido, Japan	-
Yponomeuta evonymella	Yponomeutidae	Ondrasov, Czech Republic	*Prunus padus*
Yponomeuta orientalis	Yponomeutidae	Takizawa, Iwate, Japan	-
Yponomeuta cagnagella	Yponomeutidae	Amsterdam, Netherlands	*Euonymus europaeus*
Yponomeuta padella	Yponomeutidae	Malden, Netherlands	*Prunus spinoza*
Yponomeuta malinella	Yponomeutidae	Arnhem, Netherlands	*Malus sp.*

* details in Shelton AM, Cooley RJ, Kroening MK et al (1991) Comparative analysis of two rearing procedures for diamondback moth (Lepidoptera: Plutellidae). *J Entomol Sci* 26:17–26

Supplementary Table S2: List of primers used for screening of BAC library of *Y. evonymella*.

Gene	Forward primer	Reverse primer
Ace2	AGCTGGAGCTGTTTCTGTCTC	CGCATAATGCTCTCTTCTCTTG
Henna	AACCTCAGCCACATCGAGTC	GCTCTGAGCCATACGACAGG
Kettin	CGCGCGTAAATGTAGTCCAC	TAGGCGATTCCACCATGAGG
Arp6	CGAGACTCCGCTGATAGTCAC	TTCCTGTAGTTGGACTCGGC
Plep1	AGTTCTTGCAGTCCACCTCG	GACGACTTCCTCTGTGCCAA

Supplementary Table S3: List of primers used in qPCR experiments.

Species	Gene	Forward primer	Reverse primer	Primer concentration
Y. evonymella	*Ace2*	AGCTGGAGCTGTTTCTGTCTC	CGCATAATGCTCTCTTCTCTTG	0.8 µM
	Henna	ACCAGAGTCGAAGTGCCATC	GTTCATGGTGGAATGCGAGC	0.8 µM
	Kettin	CGCGCGTAAATGTAGTCCAC	TAGGCGATTCCACCATGAGG	0.8 µM
	Arp6	CGAGACTCCGCTGATAGTCAC	TTCCTGTAGTTGGACTCGGC	0.8 µM
	Plep1	AGTTCTTGCAGTCCACCTCG	GACGACTTCCTCTGTGCCAA	0.8 µM
Y. tokyonella	*Ace2*	AGCTGGAGCTGTTTCTGTCTC	CGCATAATGCTCTCTTCTCTTG	0.4 µM
	Henna	CCGGCTACGAGTTCATGGTC	TAATGTTGAGATAGCCCCCGTC	0.4 µM
	Kettin	CGCGCGTAAATGTAGTCCAC	TAGGCGATTCCACCATGAGG	0.4 µM
	Arp6	CGAGACTCCGCTGATAGTCAC	TTCCTGTAGTTGGACTCGGC	0.4 µM
	Plep1	AGTTCTTGCAGTCCACCTCG	GACGACTTCCTCTGTGCCAA	0.4 µM
Y. plumbella	*Ace2*	AGCTGGAGCTGTTTCTGTCTC	CGCATAATGCTCTCTTCTCTTG	0.4 µM
	Henna	ACCAGAGTCGAAGTGCCATC	GTTCATGGTGGAATGCGAGC	0.4 µM
	Kettin	CGCGCGTAAATGTAGTCCAC	TAGGCGATTCCACCATGAGG	0.4 µM
	Arp6	CGAGACTCCGCTGATAGTCAC	CGAGACTCCGCTGATAGTCAC	0.4 µM
	Plep1	GGAGACGAACTGGCAGATGAA	CGACTTCCTCTGTGCCAACTA	0.4 µM

Supplementary Table S4: Summary of qPCR results. Target gene to reference ratio (*R*) was determined in three biological samples (I-III) in males (M) and females (F) using the reaction efficiencies $E_{Reference}$ and E_{Target}. The mean and its standard error (S.E.) was calculated from these three independent R values. Two null hypotheses were tested by unpaired two-tailed t-test for unequal variances. In the autosomal hypothesis (A) we tested female-to-male R ratio 1:1, whereas in the Z-linkage hypothesis (Z) the tested female-to-male ratio was 1:2. P-value lower than 0.05 means significant difference from tested ratio.

			Target gene to reference ratio (*R*)						*P* value of *t*-test	
Species	**Target gene**	**Sex**	**sample I**	**sample II**	**sample III**	$E_{Reference}$	E_{Target}	**mean ± S.D.**	**A**	**Z**
Yponomeuta plumbella	*Plep1*	F	0,182	0,175	0,181	0,877	0,905	0,179±0,004	0,0006	0,406
		M	0,375	0,333	0,321			0,343±0,028		
	Arp6	F	0,075	0,086	0,083	0,877	1,013	0,082±0,006	0,0004	0,1005
		M	0,138	0,155	0,143			0,146±0,009		
	Kettin	F	0,178	0,139	0,179	0,877	0,922	0,165±0,022	0,0004	0,441
		M	0,331	0,372	0,362			0,355±0,022		
	Henna	F	0,075	0,077	0,065	0,877	0,91	0,072±0,007	0,121	0,001
		M	0,083	0,077	0,087			0,082±0,005		
Yponomeuta tokyonella	*Plep1*	F	0,335	0,383	0,348	1,015	1,008	0,355±0,025	0,009	0,888
		M	0,739	0,559	0,798			0,699±0,125		
	Arp6	F	0,699	0,565	0,918	1,015	0,998	0,727±0,178	0,008	0,411
		M	1,199	1,249	1,337			1,262±0,07		
	Kettin	F	0,468	0,541	0,541	1,015	0,951	0,516±0,042	0,023	0,764
		M	0,873	0,993	1,39			1,085±0,271		
	Henna	F	0,455	0,396	0,312	1,015	0,999	0,388±0,072	0,0004	0,299
		M	0,855	0,902	0,871			0,876±0,024		
Yponomeuta evonymella	*Plep1*	F	0,245	0,223	0,217	0,89	0,96	0,228±0,015	0,002	0,224
		M	0,364	0,45	0,423			0,412±0,044		
	Arp6	F	0,554	0,568	0,511	0,89	0,903	0,545±0,029	0,003	0,196
		M	1,439	1, 06	1,31			1,269±0,193		
	Kettin	F	0,13	0,114	0,158	0,89	0,996	0,134±0,022	0,012	0,598
		M	0,205	0,284	0,255			0,248±0,039		
	Henna	F	4,179	3,526	4,046	0,89	0,744	3,917±0,345	0,002	0,313
		M	7,646	6,321	7,525			7,164±0,732		

4. Synthesis and perspectives

The role of CRs in shaping evolution, and their connection to speciation, has been studied for many years. Both models and empirical research have focused mainly on monocentric chromosomes even though holocentric chromosomes are relatively common, and their unique features could provide novel insights into the understanding of evolutionary processes (Lucek et al., 2022). The order Lepidoptera represents the most specious group with holocentric chromosomes. However, as lepidopteran chromosomes are generally small and undistinguishable from one another, comparative analyses have been limited to species with available genetic and physical maps and chromosome-level assemblies (e.g. Ahola et al., 2014; Hill et al., 2019). Until recently, such techniques were still relatively costly and time-consuming, and research of non-model species was limited to comparative analyses of chromosomes bearing various cytogenetic markers such as repetitive sequences, gene families, or sex chromosomes (Nguyen et al., 2010; Šíchová et al., 2015; Cabral-de-Mello et al., 2021). We explored karyotype evolution of lepidopteran species and tried to better understand the mechanism and forces behind their CRs. We combined cytogenetic techniques and bioinformatic tools to study non-model species across the whole order Lepidoptera, with a focus on non-ditrysian and early diverging ditrysian groups, which are generally understudied.

4.1. Gene families and repetitive sequences

Mapping of universal cytogenetic markers, i.e. gene families such as rDNA, histones, or snRNA genes (Provazníková et al., 2021) have been widely used in various evolutionary studies of Eukaryotes (e.g. Cai et al., 2006; Cabral-de-Mello et al., 2010; Cabral-de-Mello et al., 2011). They are easy to detect and do not require prior genetic knowledge of species under study as they are highly conserved across species (Cabral-de-Mello et al., 2011). Mapping of histone H3 genes performed by Provazníková et al. (2021) Chapter I showed that the histone cluster organization appears to reflect the overall stability of karyotype evolution across the whole order Lepidoptera, which is in agreement with previous reports in Lepidoptera (Šíchová et al., 2013, 2015) and many other animal taxa (Cavalcante et al., 2018; García-Souto et al., 2018; Zattera et al., 2020). However, the 5S rDNA and U-rich snRNA genes proved to be unsuitable markers in Lepidoptera as no clear hybridization signals were observed (Chapter I). To learn more about distribution patterns of these gene families in Lepidoptera, assemblies based on long sequencing reads would allow identification of spacer sequences between individual genes and their associations,and could reveal causes of their mobilisation. For example, in all studied species of freshwater fish of the genus *Triportheus*, U1 rRNA units were found to be present in the non-transcribed spacers (NTS) of 5S rRNA genes (Yano et al., 2020). Moreover, different types of non-LTR and LTR retrotransposons and DNA transposons were also found in 5S rDNA NTS of several species and were proposed to mediate evolution of the 5S/U1

arrays. Interestingly, U1 rRNA and 5S rRNA genes could also serve as a substrate for evolution of new SINEs (Kojima, 2015).

The most frequently used cytogenetic markers are major rRNA genes. There are databases of hybridization patterns available for animals (Sochorová et al., 2018) and plants (Garcia et al., 2012). The numbers of major rDNA loci can vary from one to 54 clusters per diploid animal genome. The majority of animal species have two clusters located near termini of chromosomes within diploid genomes, which shows that multiple clusters are relatively rare (Sochorová et al., 2018 and references within). Our results (Provazníková et al., 2021) Chapter I show that a similar pattern is also observed in the order Lepidoptera, although changes in major rDNA distribution seem to be very dynamic, and contrast with their relatively stable chromosome numbers (Robinson, 1971). These findings indicate that differences in number and position of major rDNA clusters between species do not necessarily reflect large-scale CRs.

Our analysis of long and short sequencing reads, combined with FISH, revealed the presence of complete R retrotransposons in the 28S rDNA genes, or various incomplete MEs in the IGS region of major rDNA clusters. However, as discussed in Chapter II, these MEs probably do not contribute to rDNA repatterning. Besides MEs, we also revealed the presence of satellite sequences in IGS regions in several species under study. For example, a microsatellite sequence corresponding to the insect telomere motif $(TTAGG)_n$ was found inserted into the IGS regions of *Phymatopus californicus* (Hepialidae), which have two terminal major rDNA clusters per haploid genome. Telomeric sequences are usually present in chromosome termini but can be found also elsewhere in the genome and can be responsible for various mutations and CRs (Ruiz-Herrera et al., 2008; Aksenova et al., 2013). In case of nymphalids with multiple major rDNA clusters, satellite sequences were found in their IGS regions but also elsewhere in their genomes (Dalíková et al., 2022, Chapter II). Therefore, it is possible that major rDNA units could spread to new genomic loci by ectopic recombination between repetitive sequences or via extrachromosomal circular DNA (eccDNA) (Nguyen et al., 2010; Dalíková et al., 2022). Repeats dispersed throughout the genome then serve as a substrate for recombination or insertion of eccDNA, leading to spreading of rDNA genes (Cohen et al., 2010). Interestingly, eccDNAs are known to preferentially reintegrate into telomeric regions, which could establish new rDNA loci, and upon extrachromosomal amplification could cause an increase in rDNA units in the genome (van Leen et al., 2022). The presence of eccDNAs in lepidopteran genomes could be studied using novel bioinformatics pipelines designed for NGS datasets (Mann et al., 2022).

The efficiency of ectopic recombination depends on spatial proximity of homologous sequences (Goldman and Lichten, 1996; Nguyen et al., 2010). For instance, during meiotic prophase in apes, chromosomal ends form several clusters localized to the nuclear periphery, but their clustering is determined by sequence content in the terminal chromosomal regions. If chromosomal ends contain specific satellite sequences, they never cluster with chromosomes bearing a terminal rDNA cluster and thus prevent dispersion of rDNA to these chromosomes. However, rDNA-bearing chromosomes cluster together with chromosomes with terminal heterochromatin blocks (reviewed in Hirai et al., 2019). During interphase, chromosomes are decondensed and occupy discreet regions called chromosomal territories (CTs), which are organised in nuclei based on several characteristics differing between cell types, tissues, and species. For example, in human lymphocytes, larger and gene-rich chromosomes with highly active transcription tend to be located in the centre of nuclei, with gene-poor and heterochromatic chromosomes on the periphery, whereas in human fibroblasts, size-dependent distribution is observed (reviewed in Fritz et al., 2019). Interestingly, CTs bearing NORs (nucleolus organizing regions) can interact with each other as they form nucleoli (Pliss et al., 2015). However, CTs have been described only in few non-mammalian eukaryotes (Zhang and Wang, 2021; Lukyanchikova et al., 2022). In dipterans, there is no evidence for CTs and therefore they don't seem to play a major regulatory role in the spatial organisation of chromosomes or in inter-chromosomal interactions (Kaushal et al., 2021; Lukyanchikova et al., 2022). However, for example in *Anopheles* mosquitoes, gene-rich chromatin and regions of heterochromatin from multiple chromosomes cluster together based on chromatin type, rather than chromosome identity. Unfortunately, no information about spatial distribution of chromosomes during interphase and prophase is available for lepidopterans, however there is some data on chromosome arrangement in metaphase (Lukhtanov and Dantchenko, 2002; Lukhtanov, 2019). During metaphase, chromosomes or chromosomal bivalents are ordered by size along the metaphase plate. However, this pattern is not conserved across Lepidoptera, as different metaphase chromosome spatial arrangements were found in several species, which could be caused by the holocentric nature of their chromosomes (Lukhtanov, 2019) or by specific microtubule attachment patterns (Dutrillaux et al., 2022). Nevertheless, if chromosomal configuration in Lepidoptera follows a spatial distribution pattern similar to apes or *Anopheles* mosquitoes, open chromatin of different NORs and their intermingling within nucleoli could lead to interchromosomal rearrangements (Branco and Pombo, 2006), causing sequence homogenisation, and the resulting association with heterochromatin of other chromosomes could lead to a spread of rRNA genes, satellite sequences, or both (Hirai, 2020). Modern techniques such as Hi-C sequencing (Stevens et al., 2017), super resolution microscopy (Hao et al., 2021) or serial block-face scanning electron microscopy (SBFSEM) (Yusuf et al., 2022) could

reveal chromosome spatial distribution and 3D ultrastructure and facilitate an understanding of chromatin interactions during interphase and prophase in Lepidoptera.

The association of satellite sequences and major rDNA, discovered by Dalíková et al. (2022) (Chapter II) clearly showed that despite their low abundance, satellite sequences can significantly contribute to genome reorganization in Lepidoptera. Yet, we have just hit the tip of the iceberg. To understand better the impact of satellite insertions on the major rDNA evolution, spread, and transcription, further research is needed. Analysis of assembled long read sequences could reveal distribution and localization of the satellite sequences in the rDNA cluster and/or elsewhere in the genome, and could be verified by a fibre-FISH with rDNA probes and satellite probes (de Souza et al., 2020). Such analyses could be carried out on specimens from different populations to see how stable or variable the associations are. Moreover, we show that examining closely related species, with both derived and conserved major rDNA patterns, could reveal incidence and evolutionary trends in association between satellites and rDNA.

Research of satellite sequences is challenging and very little is known about them in Trichoptera and Lepidoptera. Satellite sequences represent a very small portion of their genomes (Hebert et al., 2019; Olsen et al., 2021; Heckenhauer et al., 2022), and as some available data show, satellite sequences are scattered throughout the genomes but can be enriched on the W chromosomes or be W-specific (Dalíková et al., 2017b; Cabral-de-Mello et al., 2021). For example, satellite sequences represent only 0.255% of the genome in Crambidae moths and even less (0.01%) in *L. dispar* (Sparks et al., 2021). Information about satellites in Lepidoptera were mostly obtained by molecular techniques and analyses of genomes assembled from short reads, which provide only limited information about repeat content. In the case of short read sequencing technology, repetitive sequences can have incorrect length, often being underrepresented or even missing in the final dataset due to difficulties during library preparation (such as bias against GC-rich sequences during PCR amplification) or repetitive sequence misassembly (Marz et al., 2008; Alkan et al., 2011; Lower et al., 2018). Therefore, when designing studies focused on repetitive sequences, various methods and strategies, and their limitations, should be considered.

A combination of long read sequencing technologies and cytogenetic techniques seems to be an appropriate toolkit to disentangle the complexity of repeat structures and facilitate research in this field (Vondrak et al., 2021). Additionally, complete genomic and epigenetic maps of centromeres, which have been missing from the human genome sequences for decades, have recently been finally generated due to implementation of cutting edge technologies such as native chromatin immunoprecipitation (NChIP-seq), CUT&RUN, and assembly from highly accurate long-

read sequencing data (Altemose et al., 2022). Potentially such tools and techniques could overcome the difficulties associated with even larger repetitive regions, identify higher-order repeat structures, and help to generate high quality assemblies of heterochromatic regions and even whole chromosomes, such as the lepidopteran W chromosome (Wan et al., 2019; Lewis et al., 2021).

4.2. Sex chromosomes

Due to their many interesting features, genetic composition, and significant role in many biological processes, sex chromosomes are a popular subject of research and therefore the most studied elements in the lepidopteran karyotypes (e.g. Dalíková et al., 2017; Fraïsse et al., 2017; Yoshido et al., 2020). Recently, there has been a growing list of studies which have confirmed that the large sex chromosomes in these species result from fusion of sex chromosomes and autosomes forming neo-sex chromosomes (e.g. Nguyen et al., 2013; Dalíková et al., 2017; Mongue et al., 2017; Hill et al., 2019; Steward et al., 2021)

Our results from Chapters III and IV contribute to this list, adding two more taxa with neo-sex chromosomes which occurred independently in the superfamily Gelechioidea (Carabajal Paladino et al., 2019) and genus *Yponomeuta* (Chapter IV, Provazníková et al. 2022). The fusion of a sex chromosome in Gelechioidea with an autosome corresponding to chromosome 7 in *B. betularia* is hypothesised to be driven by ecological adaptation as the originally autosomal "performance genes" became Z-linked. Moreover, this fusion event could further lead to ecological speciation, which is typical for phytophagous insect (Jousselin and Elias, 2019), and significantly contribute to diversification (Carabajal Paladino et al., 2019).

In the genus *Yponomeuta* (Yponomeutidae), we showed that the sex chromosome trivalent in females consists of the neo-W chromosome, which pairs with the Z_1 chromosome and the Z_2 chromosome. The Z_2 chromosome was identified as an autosome corresponding to chromosome 2 in the *B. mori* reference genome (Bombycidae) (Yoshido et al., 2005a). According to Ahola et al. (2014), small chromosomes contain more repetitive DNA and are more prone to CRs. This could be the case with chromosome 2, which is one of the smallest elements in *B. mori* karyotype (Yoshido et al., 2005a). Since small chromosomes have low gene density, the fusion does not have such deleterious effect and can get fixed in the population (Anderson et al., 2020). Indeed, the synteny block corresponding to the *B. mori* chromosome 2 was also found to be fused with various chromosomes in *S. cynthia* (2n = 25-28) (Yoshido et al., 2011), *Manduca sexta* (n=28) (Yasukochi et al., 2009), and two *Pieris* species (Hill et al., 2019; Steward et al., 2021).

Alternatively, the fusion event in yponomeutids could be driven by sexual antagonism as in *B. mori,* where the majority of ovary-specific genes cluster on several autosomes, including

chromosome 2 (Suetsugu et al., 2013). Thus, the major cluster of ovary-specific genes could have become W-linked upon fusion between the W chromosome and chromosome 2 in a common ancestor of *Yponomeuta* and *Teinoptilla* genera. Linkage of ovary-specific genes to the W chromosome could have conferred short-term advantage, driving fixation of the fusion, while in the long term it represented an evolutionary trap (Anderson et al., 2020). Similar findings were discovered in three-spined stickleback fish, where the S-A fusions are also explained by sexual antagonism (Kitano et al., 2009; Dagilis et al., 2022). Yet, sexual antagonism as a driving force of sex chromosome differentiation and S-A fusions does not have strong empirical evidence in nature (Palmer et al., 2019; Perrin, 2021; Dagilis et al., 2022). Meiotic sex chromosome inactivation and incomplete dosage compensation have hampered efforts to study genomic distribution of sexually antagonistic genes in various taxa. However, that is not the case in Lepidoptera. Due to the absence of meiotic sex chromosome inactivation (Dalíková, 2017), fully balanced expression of sex-linked genes between sexes (Gu et al., 2017), and frequent turnover of sex chromosomes in some taxa (e.g. Mongue et al., 2017; Carabajal Paladino et al., 2019; Hill et al., 2019), Lepidoptera represent a great model to further investigate the potential role of sexual antagonism as a driving force of S-A fusions. Unfortunately, a direct detection of sexual antagonism via increased F_{ST} in pseudoautosomal regions (Kirkpatrick and Guerrero, 2014; Dagilis et al., 2022) is not possible in Lepidoptera due to the lack of recombination in female meiosis (Marec and Traut, 1993). Instead, using transcriptomic data from male and female early instar larvae and imagoes could identify sexually antagonistic genes (Mank and Ellegren, 2009; Dutoit et al., 2018). Subsequent analysis of distribution of these genes in the genome could test for their enrichment on the (neo-) sex chromosomes (Zhou and Bachtrog, 2012). Comparison between ermine moths with neo-sex chromosomes and, for example, the diamondback moth *Plutella xylostella* (Yponomeutoidea) with an ancestral sex chromosome system (Ward et al., 2021) could reveal whether the fused autosome was enriched for sexually antagonistic genes prior to the S-A fusion.

Apart from a few taxa (Lukhtanov, 2015; Šíchová et al., 2015), lepidopteran karyotypes are considered to be relatively stable (Robinson, 1971), which can probably be explained by a low portion of repetitive sequences which normally increase the probability of chromosomal breaks in larger chromosomes (Lavoie et al., 2013). Holocentric chromosomes alleviate some of the fitness costs of chromosomal breaks as fusion and fission products can be successfully inherited, in contrast to monocentric chromosomes (reviewed in Mandrioli and Manicardi, 2020). However, it seems that telomeres and other satellite repeats can shape lepidopteran genomes, including sex chromosomes, and therefore deserve our further attention. Furthermore, we show that CRs involving sex chromosomes such as S-A fusions are probably not random and are possibly favoured by sexually

antagonistic selection in Lepidoptera (cf. Anderson et al., 2020). The results presented in this thesis thus contribute to the ongoing efforts to elucidate the drivers of CRs and their role in evolution of Lepidoptera.

5. References

Aguilera, P. M., Debat, H. J., Scaldaferro, M. A., Martí, D. A., and Grabiele, M. (2016). FISH-mapping of the 5S rDNA locus in chili peppers (Capsicum-Solanaceae). *An Acad Bras Ciênc* 88, 117–125. doi: 10.1590/0001-37652301620140616.

Ahola, V., Lehtonen, R., Somervuo, P., Salmela, L., Koskinen, P., Rastas, P., et al. (2014). The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat Commun* 5, 4737. doi: 10.1038/ncomms5737.

Aksenova, A. Y., Greenwell, P. W., Dominska, M., Shishkin, A. A., Kim, J. C., Petes, T. D., et al. (2013). Genome rearrangements caused by interstitial telomeric sequences in yeast. *Proc Natl Acad Sci USA* 110, 19866–19871. doi: 10.1073/pnas.1319313110.

Alkan, C., Sajjadian, S., and Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nat Methods* 8, 61–65. doi: 10.1038/nmeth.1527.

Altemose, N., Logsdon, G. A., Bzikadze, A. V., Sidhwani, P., Langley, S. A., Caldas, G. V., et al. (2022). Complete genomic and epigenetic maps of human centromeres. *Science* 376, 6588. doi: 10.1126/science.abl4178.

Anderson, N. W., Hjelmen, C. E., and Blackmon, H. (2020). The probability of fusions joining sex chromosomes and autosomes. *Biol Lett* 16, 20200648. doi: 10.1098/rsbl.2020.0648.

Azevedo, N. F., Svartman, M., Manchester, A., de Moraes-Barros, N., Stanyon, R., and Vianna-Morgante, A. M. (2012). Chromosome painting in three-toed sloths: a cytogenetic signature and ancestral karyotype for Xenarthra. *BMC Evol Biol* 12, 36. doi: 10.1186/1471-2148-12-36.

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* 14, 113–124. doi: 10.1038/nrg3366.

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., et al. (2014). Sex Determination: Why so many ways of doing It? *PLoS Biol* 12, e1001899. doi: 10.1371/journal.pbio.1001899.

Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F., and Yant, L. (2018). The “polyploid hop”: shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front Ecol Evol* 6, 117. doi: 10.3389/fevo.2018.00117.

Bardell, D. (1994). Some Ancient-Greek ideas on evolution. *Am Biol Teach* 56, 198–200. doi: 10.2307/4449795.

Barton, N. H., and de Cara, M. A. R. (2009). The evolution of strong reproductive isolation. *Evolution* 63, 1171–1190. doi: 10.1111/j.1558-5646.2009.00622.x.

Bedo, D. G. (1984). Karyotypic and chromosome banding studies of the potato tuber moth, *Phthorimaea operculella* (Zeller) (Lepidoptera, Gelechiidae). *Can J Genet Cytol* 26, 141–145. doi: 10.1139/g84-024.

Beldade, P., Saenko, S. V., Pul, N., and Long, A. D. (2009). A gene-based linkage map for *Bicyclus anynana* butterflies allows for a comprehensive analysis of synteny with the lepidopteran reference genome. *PLoS Genet* 5, e1000366. doi: 10.1371/journal.pgen.1000366.

Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev* 23, 781–783. doi: 10.1101/gad.1787609.

Blackmon, H., Ross, L., and Bachtrog, D. (2017). Sex determination, sex chromosomes, and karyotype evolution in insects. *J Heredity* 108, 78–93. doi: 10.1093/jhered/esw047.

Bolnick, D. I., and Fitzpatrick, B. M. (2007). Sympatric speciation: models and empirical evidence. *Annu Rev Ecol Evol Syst* 38, 459–487. doi: 10.1146/annurev.ecolsys.38.091206.095804.

Bracewell, R., Chatla, K., Nalley, M. J., and Bachtrog, D. (2019). Dynamic turnover of centromeres drives karyotype evolution in Drosophila. *eLife* 8, e49002. doi: 10.7554/eLife.49002.

Branco, M. R., and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* 4, e138. doi: 10.1371/journal.pbio.0040138.

Brooke, B. D., Hunt, R. H., Chandre, F., Carnevale, P., and Coetzee, M. (2002). Stable chromosomal inversion polymorphisms and insecticide resistance in the malaria vector mosquito *Anopheles gambiae* (Diptera: Culicidae). *J Med Entomol* 39, 568–573. doi: 10.1603/0022-2585-39.4.568.

Brown Jr, K. S., Von Schoultz, B., and Suomalainen, E. (2005). Chromosome evolution in Neotropical Danainae and Ithomiinae (Lepidoptera): Chromosome evolution in Neotropical Danainae and Ithomiinae. *Hereditas* 141, 216–236. doi: 10.1111/j.1601-5223.2004.01868.x.

Cabral-de-Mello, D. C., Moura, R. C., and Martins, C. (2010). Chromosomal mapping of repetitive DNAs in the beetle *Dichotomius geminatus* provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA similarity between the B chromosome and A complement. *Heredity* 104, 393–400. doi: 10.1038/hdy.2009.126.

Cabral-de-Mello, D. C., Oliveira, S. G., de Moura, R. C., and Martins, C. (2011). Chromosomal organization of the 18S and 5S rRNAs and histone H3 genes in Scarabaeinae coleopterans: insights into the evolutionary dynamics of multigene families and heterochromatin. *BMC Genetics* 12, 88. doi: 10.1186/1471-2156-12-88.

Cabral-de-Mello, D. C., Zrzavá, M., Kubíčková, S., Rendón, P., and Marec, F. (2021). The role of satellite DNAs in genome architecture and sex chromosome evolution in Crambidae moths. *Front Genet* 12, 661417. doi: 10.3389/fgene.2021.661417.

Cai, Q., Zhang, D., Liu, Z.-L., and Wang, X.-R. (2006). Chromosomal localization of 5S and 18S rDNA in five species of subgenus *Strobus* and their implications for genome evolution of *Pinus*. *Ann Bot* 97, 715–722. doi: 10.1093/aob/mcl030.

Carabajal Paladino, L. Z., Ferrari, M. E., Lauría, J. P., Cagnotti, C. L., Šíchová, J., and López, S. N. (2016). The effect of X-rays on cytological traits of *Tuta absoluta* (Lepidoptera: Gelechiidae). *Fla. Entomol.* 99, 43–53. doi: 10.1653/024.099.sp107.

Carabajal Paladino, L. Z., Nguyen, P., Šíchová, J., and Marec, F. (2014). Mapping of single-copy genes by TSA-FISH in the codling moth, *Cydia pomonella*. *BMC Genet* 15, S15. doi: 10.1186/1471-2156-15-S2-S15.

Carabajal Paladino, L. Z., Provazníková, I., Berger, M., Bass, C., Aratchige, N. S., López, S. N., et al. (2019). Sex chromosome turnover in moths of the diverse superfamily Gelechioidea. *Genome Biol Evol* 11, 1307–1319. doi: 10.1093/gbe/evz075.

Cavalcante, M. G., Bastos, C. E. M. C., Nagamachi, C. Y., Pieczarka, J. C., Vicari, M. R., and Noronha, R. C. R. (2018). Physical mapping of repetitive DNA suggests 2n reduction in Amazon turtles *Podocnemis* (Testudines: Podocnemididae). *PLoS ONE* 13, e0197536. doi: 10.1371/journal.pone.0197536.

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci USA* 75, 5618–5622. doi: 10.1073/pnas.75.11.5618.

Charlesworth, B., and Charlesworth, D. (2000). The degeneration of Y chromosomes. *Phil Trans R Soc Lond B* 355, 1563–1572. doi: 10.1098/rstb.2000.0717.

Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95, 118–128. doi: 10.1038/sj.hdy.6800697.

Chi, J., Fu, B., Nie, W., Wang, J., Graphodatsky, A. S., and Yang, F. (2005). New insights into the karyotypic relationships of Chinese muntjac *(Muntiacus reevesi)*, forest musk deer *(Moschus berezovskii)* and gayal *(Bos frontalis)*. *Cytogenet Genome Res* 108, 310–316. doi: 10.1159/000081520.

Christmas, M. J., Wallberg, A., Bunikis, I., Olsson, A., Wallerman, O., and Webster, M. T. (2019). Chromosomal inversions associated with environmental adaptation in honeybees. *Mol Ecol* 28, 1358–1374. doi: 10.1111/mec.14944.

Cicconardi, F., Lewis, J. J., Martin, S. H., Reed, R. D., Danko, C. G., and Montgomery, S. H. (2021). Chromosome fusion affects genetic diversity and evolutionary turnover of functional loci but consistently depends on chromosome size. *Mol Biol Evol* 38, 4449–4462. doi: 10.1093/molbev/msab185.

Clark, F. E., and Kocher, T. D. (2019). Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. *Sci Rep* 9, 20213. doi: 10.1038/s41598-019-55774-8.

Cohen, S., Agmon, N., Sobol, O., and Segal, D. (2010). Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mobile DNA* 1, 11. doi: 10.1186/1759-8753-1-11.

Conte, M. A., Joshi, R., Moore, E. C., Nandamuri, S. P., Gammerdinger, W. J., Roberts, R. B., et al. (2019). Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *GigaScience* 8, 1–20. doi: 10.1093/gigascience/giz030.

Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., et al. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature* 508, 488–493. doi: 10.1038/nature13151.

Crosland, M. W. J., and Crozier, R. H. (1986). *Myrmecia pilosula*, an ant with only one pair of chromosomes. *Science* 231, 1278. doi: 10.1126/science.231.4743.1278.

Dagilis, A. J., Sardell, J. M., Josephson, M. P., Su, Y., Kirkpatrick, M., and Peichel, C. L. (2022). Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes. *Phil Trans R Soc B* 377, 20210205. doi: 10.1098/rstb.2021.0205.

Dalíková, M., Provazníková, I., Provazník, J., Grof-Tisza, P., Pepi, A., and Nguyen, P. (2022). The role of repetitive DNA in re-patterning of major rDNA clusters in Lepidoptera. *bioRxiv*. doi: 10.1101/2022.03.26.485928.

Dalíková, M., Zrzavá, M., Hladová, I., Nguyen, P., Šonský, I., Flegrová, M., et al. (2017a). New insights into the evolution of the W chromosome in Lepidoptera. *J Hered* 108, 709–719. doi: 10.1093/jhered/esx063.

Dalíková, M., Zrzavá, M., Kubíčková, S., and Marec, F. (2017b). W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). *Chromosome Res* 25, 241–252. doi: 10.1007/s10577-017-9558-8.

De Prins, J., and Saitoh, K. (2003). "Karyology and sex determination.," in *Lepidoptera, moths and butterflies: morphology, physiology, and development* (Berlin, Germany: Walter de Gruyter), 449–468.

de Souza, T. B., Gaeta, M. L., Martins, C., and Vanzela, A. L. L. (2020). IGS sequences in *Cestrum* present AT- and GC-rich conserved domains, with strong regulatory potential for 5S rDNA. *Mol Biol Rep* 47, 55–66. doi: 10.1007/s11033-019-05104-y.

de Vos, J. M., Augustijnen, H., Bätscher, L., and Lucek, K. (2020). Speciation through chromosomal fusion and fission in Lepidoptera. *Phil Trans R Soc B* 375, 20190539. doi: 10.1098/rstb.2019.0539.

Deakin, J. E., Potter, S., O'Neill, R., Ruiz-Herrera, A., Cioffi, M. B., Eldridge, M. D. B., et al. (2019). Chromosomics: bridging the gap between genomes and chromosomes. *Genes* 10, 627. doi: 10.3390/genes10080627.

Dobzhansky, T. (1950). The chromosomes of *Drosophila willistoni*. *J Hered* 41, 156–158. doi: 10.1093/oxfordjournals.jhered.a106117.

Dutoit, L., Mugal, C. F., Bolívar, P., Wang, M., Nadachowska-Brzyska, K., Smeds, L., et al. (2018). Sex-biased gene expression, sexual antagonism and levels of genetic diversity in the collared flycatcher (*Ficedula albicollis*) genome. *Mol Ecol* 27, 3572–3581. doi: 10.1111/mec.14789.

Dutrillaux, B., Dutrillaux, A.-M., McClure, M., Gèze, M., Elias, M., and Bed'hom, B. (2022). Improved basic cytogenetics challenges holocentricity of butterfly chromosomes. *bioRxiv*. doi: 10.1101/2022.03.11.484012.

Ellegren, H. (2011). Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet* 12, 157–166. doi: 10.1038/nrg2948.

Ennis, T. J. (1976). Sex chromatin and chromosomes numbers in Lepidoptera. *Can. J. Genet. Cytol.* 18, 119–130. doi: 10.1139/g76-017.

Fang, Z., Pyhäjärvi, T., Weber, A. L., Dawe, R. K., Glaubitz, J. C., González, J. de J. S., et al. (2012). Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* 191, 883–894. doi: 10.1534/genetics.112.138578.

Faria, R., and Navarro, A. (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol Evol* 25, 660–669. doi: 10.1016/j.tree.2010.07.008.

Fierst, J. L., and Hansen, T. F. (2010). Genetic architecture and postzygotic reproductive isolation: evolution of Bateson-Dobzhansky-Muller incompatibilities in a polygenic model. *Evolution* 64, 675–693. doi: 10.1111/j.1558-5646.2009.00861.x.

Fitzpatrick, B. M., Fordyce, J. A., and Gavrilets, S. (2008). What, if anything, is sympatric speciation? *J Evol Biol* 21, 1452–1459. doi: 10.1111/j.1420-9101.2008.01611.x.

Fraïsse, C., Picard, M. A. L., and Vicoso, B. (2017). The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nat Commun* 8, 1486. doi: 10.1038/s41467-017-01663-5.

Fritz, A. J., Sehgal, N., Pliss, A., Xu, J., and Berezney, R. (2019). Chromosome territories and the global regulation of the genome. *Genes Chromosom Cancer* 58, 407–426. doi: 10.1002/gcc.22732.

Fuková, I., Nguyen, P., and Marec, F. (2005). Codling moth cytogenetics: Karyotype, chromosomal location of rDNA, and molecular differentiation of sex chromosomes. *Genome* 48, 1083–1092. doi: 10.1139/g05-063.

Fuková, I., Traut, W., Vítková, M., Nguyen, P., Kubíčková, S., and Marec, F. (2007). Probing the W chromosome of the codling moth, *Cydia pomonella*, with sequences from microdissected sex chromatin. *Chromosoma* 116, 135–145. doi: 10.1007/s00412-006-0086-0.

Furman, B. L. S., Metzger, D. C. H., Darolti, I., Wright, A. E., Sandkam, B. A., Almeida, P., et al. (2020). Sex chromosome evolution: So many exceptions to the rules. *Genome Biol Evol* 12, 750–763. doi: 10.1093/gbe/evaa081.

Gammerdinger, W. J., and Kocher, T. D. (2018). Unusual diversity of sex chromosomes in African cichlid fishes. *Genes* 9, 480. doi: 10.3390/genes9100480.

Garagna, S., Page, J., and Fernandez-Donoso, R. (2014). The Robertsonian phenomenon in the house mouse: mutation, meiosis and speciation. *Chromosoma* 123, 529–544. doi: 10.1007/s00412-014-0477-6.

Garcia, S., Garnatje, T., and Kovařík, A. (2012). Plant rDNA database: ribosomal DNA loci information goes online. *Chromosoma* 121, 389–394. doi: 10.1007/s00412-012-0368-7.

García-Souto, D., Alonso-Rubido, S., Costa, D., Eirín-López, J., Rolán-Álvarez, E., Faria, R., et al. (2018). Karyotype characterization of nine Periwinkle species (Gastropoda, Littorinidae). *Genes* 9, 517. doi: 10.3390/genes9110517.

Gavrilets, S. (2014). Models of speciation: Where are we now? *J Hered* 105, 743–755. doi: 10.1093/jhered/esu045.

Gayon, J. (2016). From Mendel to epigenetics: history of genetics. *Comptes Rendus Biologies* 339, 225–230. doi: 10.1016/j.crvi.2016.05.009.

Goldman, A. S. H., and Lichten, M. (1996). The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. *Genetics* 144, 43–55. doi: 10.1093/genetics/144.1.43.

Gonçalves-Oliveira, R. C., Ximenes, A. F., Oliveira, A. R., Vasconcelos, S., Carvalho-Filho, N., Oliveira, G., et al. (2020). Genome size and chromosome number conservation contrasting with

karyotype diversity in *Hohenbergia* (Bromelioideae, Bromeliaceae). *Bot J Linn Soc* 192, 900–909. doi: 10.1093/botlinnean/boz094.

Gotter, A. L., Levine, J. D., and Reppert, S. M. (1999). Sex-linked period genes in the silkmoth, *Antheraea pernyi*: implications for circadian clock regulation and the evolution of sex chromosomes. *Cell Press* 24, 953–965. doi: 10.1016/S0896-6273(00)81042-9.

Gu, L., Walters, J. R., and Knipple, D. C. (2017). Conserved patterns of sex chromosome dosage compensation in the Lepidoptera (WZ/ZZ): insights from a moth neo-Z chromosome. *Genome Biol Evol* 9, 802–816. doi: 10.1093/gbe/evx039.

Guerrero, R. F., and Kirkpatrick, M. (2014). Local adaptation and the evolution of chromosome fusions. *Evolution* 68, 2747–2756. doi: 10.1111/evo.12481.

Hao, X., Parmar, J. J., Lelandais, B., Aristov, A., Ouyang, W., Weber, C., et al. (2021). Super-resolution visualization and modeling of human chromosomal regions reveals cohesin-dependent loop structures. *Genome Biol* 22, 150. doi: 10.1186/s13059-021-02343-w.

Hebert, F. O., Freschi, L., Blackburn, G., Béliveau, C., Dewar, K., Boyle, B., et al. (2019). Expansion of LINEs and species-specific DNA repeats drives genome expansion in Asian Gypsy Moths. *Sci Rep* 9, 16413. doi: 10.1038/s41598-019-52840-z.

Heckenhauer, J., Frandsen, P. B., Sproul, J. S., Li, Z., Paule, J., Larracuente, A. M., et al. (2022). Genome size evolution in the diverse insect order Trichoptera. *GigaScience* 11, 1–19. doi: 10.1093/gigascience/giac011.

Hejníčková, M., Dalíková, M., Potocký, P., Tammaru, T., Trehubenko, M., Kubíčková, S., et al. (2021). Degenerated, undifferentiated, rearranged, lost: high variability of sex chromosomes in Geometridae (Lepidoptera) identified by sex chromatin. *Cells* 10, 2230. doi: 10.3390/cells10092230.

Hejníčková, M., Koutecký, P., Potocký, P., Provazníková, I., Voleníková, A., Dalíková, M., et al. (2019). Absence of W Chromosome in Psychidae Moths and Implications for the Theory of Sex Chromosome Evolution in Lepidoptera. *Genes* 10, 1016. doi: 10.3390/genes10121016.

Henderson, A. S., Warburton, D., and Atwood, K. C. (1972). Location of ribosomal DNA in the human chromosome complement. *Proc Natl Acad Sci USA* 69, 3394–3398. doi: 10.1073/pnas.69.11.3394.

Hill, J., Rastas, P., Hornett, E. A., Neethiraj, R., Clark, N., Morehouse, N., et al. (2019). Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Sci Adv* 5, eaau3648. doi: 10.1126/sciadv.aau3648.

Hirai, H. (2020). Chromosome dynamics regulating genomic dispersion and alteration of nucleolus organizer regions (NORs). *Cells* 9, 971. doi: 10.3390/cells9040971.

Hirai, H., Hirai, Y., Udono, T., Matsubayashi, K., Tosi, A. J., and Koga, A. (2019). Structural variations of subterminal satellite blocks and their source mechanisms as inferred from the meiotic configurations of chimpanzee chromosome termini. *Chromosome Res* 27, 321–332. doi: 10.1007/s10577-019-09615-z.

Hoffmann, A. A., and Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol Syst* 39, 21–42. doi: 10.1146/annurev.ecolsys.39.110707.173532.

Huang, Z., De O. Furo, I., Liu, J., Peona, V., Gomes, A. J. B., Cen, W., et al. (2022). Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. *Nat Commun* 13, 944. doi: 10.1038/s41467-022-28585-1.

Imarazene, B., Du, K., Beille, S., Jouanno, E., Feron, R., Pan, Q., et al. (2021). A supernumerary "B-sex" chromosome drives male sex determination in the Pachón cavefish, *Astyanax mexicanus*. *Current Biol* 31, 4800-4809.e9. doi: 10.1016/j.cub.2021.08.030.

Jay, P., Tezenas, E., Véber, A., and Giraud, T. (2022). Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. *PLoS Biol* 20, e3001698. doi: 10.1371/journal.pbio.3001698.

Jeffries, D. L., Gerchen, J. F., Scharmann, M., and Pannell, J. R. (2021). A neutral model for the loss of recombination on sex chromosomes. *Phil Trans R Soc B* 376, 20200096. doi: 10.1098/rstb.2020.0096.

Jousselin, E., and Elias, M. (2019). Testing host-plant driven speciation in phytophagous insects: a phylogenetic perspective. *arXiv*. doi: 10.20944/preprints201902.0215v2.

Kaiser, V. B., and Bachtrog, D. (2010). Evolution of sex chromosomes in insects. *Annu Rev Genet* 44, 91–112. doi: 10.1146/annurev-genet-102209-163600.

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., et al. (2012). A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genet* 8, e1002798. doi: 10.1371/journal.pgen.1002798.

Kandul, N. P., Lukhtanov, V. A., and Pierce, N. E. (2007). Karyotipic diversity and speciation in *Agrodiaetus* butterflies. *Evolution* 61, 546–559. doi: 10.1111/j.1558-5646.2007.00046.x.

Kaushal, A., Mohana, G., Dorier, J., Özdemir, I., Omer, A., Cousin, P., et al. (2021). CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions. *Nat Commun* 12, 1011. doi: 10.1038/s41467-021-21366-2.

Kautt, A. F., Kratochwil, C. F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., et al. (2020). Contrasting signatures of genomic divergence during sympatric speciation. *Nature* 588, 106–111. doi: 10.1038/s41586-020-2845-0.

Khandelwal, S. (1990). Chromosome evolution in the genus *Ophioglossum* L. *Bot Jour Linn Soc* 102, 205–217. doi: 10.1111/j.1095-8339.1990.tb01876.x.

Kirkpatrick, M., and Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics* 173, 419–434. doi: 10.1534/genetics.105.047985.

Kirkpatrick, M., and Guerrero, R. F. (2014). Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics* 197, 531–541. doi: 10.1534/genetics.113.156026.

Kitano, J., and Peichel, C. L. (2012). Turnover of sex chromosomes and speciation in fishes. *Environ Biol Fish* 94, 549–558. doi: 10.1007/s10641-011-9853-8.

Kitano, J., Ross, J. A., Mori, S., Kume, M., Jones, F. C., Chan, Y. F., et al. (2009). A role for a neo-sex chromosome in stickleback speciation. *Nature* 461, 1079–1083. doi: 10.1038/nature08441.

Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., et al. (2014). A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature* 509, 633–636. doi: 10.1038/nature13315.

Kobayashi, T. (2006). Strategies to maintain the stability of the ribosomal RNA gene repeats. *Genes Genet. Syst.* 81, 155–161. doi: 10.1266/ggs.81.155.

Kojima, K. K. (2015). A new class of SINEs with snRNA gene-derived heads. *Genome Biol Evol* 7, 1702–1712. doi: 10.1093/gbe/evv100.

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B* 279, 5048–5057. doi: 10.1098/rspb.2012.1108.

Lavoie, C. A., Platt, R. N., Novick, P. A., Counterman, B. A., and Ray, D. A. (2013). Transposable element evolution in *Heliconius* suggests genome diversity within Lepidoptera. *Mobile DNA* 4, 21. doi: 10.1186/1759-8753-4-21.

Leaché, A. D., Banbury, B. L., Linkem, C. W., and de Oca, A. N.-M. (2016). Phylogenomics of a rapid radiation: is chromosomal evolution linked to increased diversification in North American spiny lizards (Genus Sceloporus)? *BMC Evol Biol* 16. doi: 10.1186/s12862-016-0628-x.

Leclercq, S., Thézé, J., Chebbi, M. A., Giraud, I., Moumen, B., Ernenwein, L., et al. (2016). Birth of a W sex chromosome by horizontal transfer of *Wolbachia* bacterial symbiont genome. *Proc Natl Acad Sci USA* 113, 15036–15041. doi: 10.1073/pnas.1608979113.

Lewis, J. J., Cicconardi, F., Martin, S. H., Reed, R. D., Danko, C. G., and Montgomery, S. H. (2021). The *Dryas iulia* genome supports multiple gains of a W chromosome from a B chromosome in butterflies. *Genome Biol Evol* 13, evab128. doi: 10.1093/gbe/evab128.

Li Causi, P. E. (2008). Hybridization as speciation? The viewpoint of Greek folk biology (and Aristotle) on the mutation of species. *SSRN Journal*. doi: 10.2139/ssrn.1323052.

Liu, Z., Roesti, M., Marques, D., Hiltbrunner, M., Saladin, V., and Peichel, C. L. (2022). Chromosomal fusions facilitate adaptation to divergent environments in threespine stickleback. *Mol Biol Evol* 39, msab358. doi: 10.1093/molbev/msab358.

Lorch, S., Zeuss, D., Brandl, R., and Brändle, M. (2016). Chromosome numbers in three species groups of freshwater flatworms increase with increasing latitude. *Ecol Evol* 6, 1420–1429. doi: 10.1002/ece3.1969.

Lower, S. S., McGurk, M. P., Clark, A. G., and Barbash, D. A. (2018). Satellite DNA evolution: old ideas, new approaches. *Curr Opin Genet Dev* 49, 70–78. doi: 10.1016/j.gde.2018.03.003.

Lowry, D. B., and Willis, J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol* 8, e1000500. doi: 10.1371/journal.pbio.1000500.

Lucek, K., Augustijnen, H., and Escudero, M. (2022). A holocentric twist to chromosomal speciation? *Trends Ecol Evol* 37, 655–662. doi: 10.1016/j.tree.2022.04.002.

Lukhtanov, V. (2015). The blue butterfly *Polyommatus (Plebicula) atlanticus* (Lepidoptera, Lycaenidae) holds the record of the highest number of chromosomes in the non-polyploid eukaryotic organisms. *Comparat Cytogenet* 9, 683–690. doi: 10.3897/CompCytogen.v9i4.5760.

Lukhtanov, V. A. (2000). Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta). *J Zool Syst Evol Res* 38, 73–79. doi: 10.1046/j.1439-0469.2000.382130.x.

Lukhtanov, V. A. (2019). Two types of highly ordered micro- and macrochromosome arrangement in metaphase plates of butterflies (Lepidoptera). *CCG* 13, 19–25. doi: 10.3897/CompCytogen.v13i1.32614.

Lukhtanov, V. A., and Dantchenko, A. V. (2002). Principles of highly ordered metaphase I bivalent arrangement in spermatocytes of *Agrodiaetus* (Lepidoptera). *Chromosome Res* 10, 5–20. doi: 10.1023/A:1014249607796.

Lukhtanov, V. A., Dantchenko, A. V., Balayan, K. V., and Gagarina, A. V. (2020). Karyotype and DNA barcode of *Polyommatus* (Agrodiaetus) *cyaneus* (Staudinger, 1899) from its type locality: implication for taxonomic and evolutionary research in *Polyommatus* blue butterflies (Lepidoptera, Lycaenidae). *CCG* 14, 567–575. doi: 10.3897/CompCytogen.v14.i4.59574.

Lukyanchikova, V., Nuriddinov, M., Belokopytova, P., Taskina, A., Liang, J., Reijnders, M. J. M. F., et al. (2022). *Anopheles* mosquitoes reveal new principles of 3D genome organization in insects. *Nat Commun* 13, 1960. doi: 10.1038/s41467-022-29599-5.

Lusinska, J., Majka, J., Betekhtin, A., Susek, K., Wolny, E., and Hasterok, R. (2018). Chromosome identification and reconstruction of evolutionary rearrangements in *Brachypodium distachyon*, *B. stacei* and *B. hybridum*. *Ann Bot* 122, 445–459. doi: 10.1093/aob/mcy086.

Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., et al. (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proc Natl Acad Sci USA* 115, E236–E243. doi: 10.1073/pnas.1713288114.

Macas, J., Neumann, P., and Navrátilová, A. (2007). Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* 8, 427. doi: 10.1186/1471-2164-8-427.

Mandrioli, M., and Manicardi, G. C. (2020). Holocentric chromosomes. *PLoS Genet* 16, e1008918. doi: 10.1371/journal.pgen.1008918.

Mank, J. E., and Avise, J. C. (2006). Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes. *Genetica* 127, 321–327. doi: 10.1007/s10709-005-5248-0.

Mank, J. E., and Ellegren, H. (2009). Sex-linkage of sexually antagonistic genes is predicted by female, but not male, effects in birds. *Evolution* 63, 1464–1472. doi: 10.1111/j.1558-5646.2009.00618.x.

Mann, L., Seibt, K. M., Weber, B., and Heitkam, T. (2022). ECCsplorer: a pipeline to detect extrachromosomal circular DNA (eccDNA) from next-generation sequencing data. *BMC Bioinformatics* 23, 40. doi: 10.1186/s12859-021-04545-2.

Marec, F., and Novak, K. (1998). Absence of sex chromatin corresponds with a sex-chromosome univalent in females of Trichoptera. *Eur J Entomol* 95, 197–209.

Marec, F., and Traut, W. (1993). Synaptonemal complexes in female and male meiotic prophase of *Ephestia kuehniella* (Lepidoptera). *Heredity* 71, 394–404. doi: 10.1038/hdy.1993.154.

Marz, M., Kirsten, T., and Stadler, P. F. (2008). Evolution of spliceosomal snRNA genes in metazoan animals. *J Mol Evol* 67, 594–607. doi: 10.1007/s00239-008-9149-6.

Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Brieuc, M. S. O., et al. (2022). Supergene origin and maintenance in Atlantic cod. *Nat Ecol Evol* 6, 469–481. doi: 10.1038/s41559-022-01661-x.

Mediouni, J., Fuková, I., Frydrychová, R., Dhouibi, M. H., and Marec, F. (2004). Karyotype, sex chromatin and sex chromosome differentiation in the carob moth, *Ectomyelois ceratoniae* (Lepidoptera: Pyralidae). *Caryologia* 57, 184–194. doi: 10.1080/00087114.2004.10589391.

Meisel, R. P. (2020). Evolution of sex determination and sex chromosomes: a novel alternative paradigm. *BioEssays* 42, 1900212. doi: 10.1002/bies.201900212.

Mendel, J. G. (1865). Versuche über Pflanzenhybriden. *Verh naturf Ver Brünn* 4, 3–37.

Mendoza, M. N., Raudsepp, T., More, M. J., Gutiérrez, G. A., and Ponce de León, F. A. (2020). Cytogenetic mapping of 35 new markers in the Alpaca (*Vicugna pacos*). *Genes* 11, 522. doi: 10.3390/genes11050522.

Mérot, C., Berdan, E. L., Cayuela, H., Djambazian, H., Ferchaud, A.-L., Laporte, M., et al. (2021). Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *Mol Biol Evol* 38, 3953–3971. doi: 10.1093/molbev/msab143.

Mita, K., Kasahara, M., Sasaki, S., Nagayashu, Y., Yamada, T., and Kanamori, H. (2004). The genome sequence of silkworm, *Bombyx mori*. *DNA Res* 11, 27–35. doi: 10.1093/dnares/11.1.27.

Mongue, A. J., Nguyen, P., Voleníková, A., and Walters, J. R. (2017). Neo-sex chromosomes in the Monarch butterfly, *Danaus plexippus*. *3G-Genes Genom Genet* 7, 3281–3294. doi: 10.1534/g3.117.300187.

Morales, H. E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A. M., et al. (2019). Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. *Sci Adv* 5, eaav9963. doi: 10.1126/sciadv.aav9963.

Morgan, T. H., Sturtevant, A. H., Muller, H. J., and Bridges, C. B. (1915). *The mechanism of Mendelian heredity*. New York: Henry Holt.

Mrackova, M., Nicolas, M., Hobza, R., Negrutiu, I., Monéger, F., Widmer, A., et al. (2008). Independent origin of sex chromosomes in two species of the genus Silene. *Genetics* 179, 1129–1133. doi: 10.1534/genetics.107.085670.

Nagaraju, J., Gopinath, G., Sharma, V., and Shukla, J. N. (2014). Lepidopteran sex determination: a cascade of surprises. *Sex Dev* 8, 104–112. doi: 10.1159/000357483.

Navarro, A., and Barton, N. H. (2003). Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science* 300, 321–324. doi: 10.1126/science.1080600.

Nguyen, P., Sahara, K., Yoshido, A., and Marec, F. (2010). Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). *Genetica* 138, 343–354. doi: 10.1007/s10709-009-9424-5.

Nguyen, P., Sýkorová, M., Šíchová, J., Kůta, V., Dalíková, M., Čapková Frydrychová, R., et al. (2013). Neo-sex chromosomes and adaptive potential in tortricid pests. *PNAS* 110, 6931–6936. doi: 10.1073/pnas.1220372110.

Nilsson, N.-O., 'Löfstedt, C., and Dävring, L. (1988). Unusual sex chromosome inheritance in six species of small ermine moths (Yponomeuta, Yponomeutidae, Lepidoptera). *Hereditas* 108, 259–265. doi: 10.1111/j.1601-5223.1988.tb00311.x.

Noor, M. A. F., Grams, K. L., Bertucci, L. A., and Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* 98, 12084–12088. doi: 10.1073/pnas.221274498.

Offner, S. (1994). Using chromosomes to teach evolution: II. chromosomal rearrangements in speciation events. *Am Biol Teach* 56, 86–93. doi: 10.2307/4449759.

Ohno, S. (1970). *Evolution by gene duplication*. 1st ed. Berlin, Germany: Springer Available at: https://doi.org/10.1007/978-3-642-86659-3.

Olsen, L. K., Heckenhauer, J., Sproul, J. S., Dikow, R. B., Gonzalez, V. L., Kweskin, M. P., et al. (2021). Draft genome assemblies and annotations of *Agrypnia vestita* Walker, and *Hesperophylax magnus* banks reveal substantial repetitive element expansion in tube case-making caddisflies (Insecta: Trichoptera). *Genome Biol Evol* 13, 1–7. doi: 10.1093/gbe/evab013.

Palmer, D. H., Rogers, T. F., Dean, R., and Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Mol Ecol* 28, 4709–4724. doi: 10.1111/mec.15245.

Pennell, M. W., Kirkpatrick, M., Otto, S. P., Vamosi, J. C., Peichel, C. L., Valenzuela, N., et al. (2015). Y Fuse? Sex chromosome fusions in fishes and reptiles. *PLoS Genet* 11, e1005237. doi: 10.1371/journal.pgen.1005237.

Perrin, N. (2021). Sex-chromosome evolution in frogs: what role for sex-antagonistic genes? *Phil Trans R Soc B* 376, 20200094. doi: 10.1098/rstb.2020.0094.

Picq, S., Lumley, L., Šíchová, J., Laroche, J., Pouliot, E., Brunet, B. M. T., et al. (2018). Insights into the structure of the spruce budworm (*Choristoneura fumiferana*) genome, as revealed by molecular cytogenetic analyses and a high-density linkage map. *G3-Genes Genom Genet* 8, 2539–2549. doi: 10.1534/g3.118.200263.

Pliss, A., Fritz, A. J., Stojkovic, B., Ding, H., Mukherjee, L., Bhattacharya, S., et al. (2015). Non-random patterns in the distribution of NOR-bearing chromosome territories in human fibroblasts: a network model of interactions: NOR-CT: a network model of interactions. *J Cell Physiol* 230, 427–439. doi: 10.1002/jcp.24726.

Pokorná, M., Altmanová, M., and Kratochvíl, L. (2014). Multiple sex chromosomes in the light of female meiotic drive in amniote vertebrates. *Chromosome Res* 22, 35–44. doi: 10.1007/s10577-014-9403-2.

Potter, S., Bragg, J. G., Blom, M. P. K., Deakin, J. E., Kirkpatrick, M., Eldridge, M. D. B., et al. (2017). Chromosomal speciation in the genomics era: disentangling phylogenetic evolution of rock-wallabies. *Front Genet* 8. doi: 10.3389/fgene.2017.00010.

Presgraves, D. C., and Meiklejohn, C. D. (2021). Hybrid sterility, genetic conflict and complex speciation: lessons from the *Drosophila simulans* clade species. *Front Genet* 12, 669045. doi: 10.3389/fgene.2021.669045.

Prokopowich, C. D., Gregory, T. R., and Crease, T. J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46, 48–50. doi: 10.1139/g02-103.

Provazníková, I., Hejníčková, M., Visser, S., Dalíková, M., Carabajal Paladino, L. Z., Zrzavá, M., et al. (2021). Large-scale comparative analysis of cytogenetic markers across Lepidoptera. *Sci Rep* 11, 12214. doi: 10.1038/s41598-021-91665-7.

Rens, W., Grützner, F., O'Brien, P. C. M., Fairclough, H., Graves, J. A. M., and Ferguson-Smith, M. A. (2004). Resolution and evolution of the duck-billed platypus karyotype with an $X_1 Y_1 X_2 Y_2 X_3 Y_3 X_4 Y_4 X_5 Y_5$ male sex chromosome constitution. *Proc Natl Acad Sci USA* 101, 16257–16261. doi: 10.1073/pnas.0405702101.

Richards, E. J., Servedio, M. R., and Martin, C. H. (2019). Searching for sympatric speciation in the genomic era. *BioEssays* 41, 1900047. doi: 10.1002/bies.201900047.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends Ecol Evol* 16, 351–358. doi: 10.1016/S0169-5347(01)02187-5.

Roberts, R. B., Ser, J. R., and Kocher, T. D. (2009). Sexual conflict resolved by invasion of a novel sex determiner in lake Malawi cichlid fishes. *Science* 326, 998–1001. doi: 10.1126/science.1174705.

Robinson, R. (1971). *Lepidoptera genetics*. Oxford: Pergamon Press.

Robinson, T. J., Cernohorska, H., Diedericks, G., Cabelova, K., Duran, A., and Matthee, C. A. (2014). Phylogeny and vicariant speciation of the Grey Rhebok, *Pelea capreolus*. *Heredity* 112, 325–332. doi: 10.1038/hdy.2013.108.

Roe, A. D., Weller, S. J., Baixeras, J., Brown, J., Cummings, M. P., and David, D. (2009). "Evolutionary Framework for Lepidoptera Model Systems," in *Genetics and Molecular Biology of Lepidoptera* (Boca Raton: CRC Press), 1–24.

Rovatsos, Altmanová, Augstenová, Mazzoleni, Velenský, and Kratochvíl (2019). ZZ/ZW sex determination with multiple neo-sex chromosomes is common in Madagascan chameleons of the genus *Furcifer* (Reptilia: Chamaeleonidae). *Genes* 10, 1020. doi: 10.3390/genes10121020.

Ruiz-Herrera, A., Nergadze, S. G., Santagostino, M., and Giulotto, E. (2008). Telomeric repeats far from the ends: mechanisms of origin and role in evolution. *Cytogenet Genome Res* 122, 219–228. doi: 10.1159/000167807.

Sahara, K., Yoshido, A., Kawamura, N., Ohnuma, A., Abe, H., Mita, K., et al. (2003). W-derived BAC probes as a new tool for identification of the W chromosome and its aberrations in *Bombyx mori*. *Chromosoma* 112, 48–55. doi: 10.1007/s00412-003-0245-5.

Sahara, K., Yoshido, A., and Traut, W. (2012). Sex chromosome evolution in moths and butterflies. *Chromosome Res* 20, 83–94. doi: 10.1007/s10577-011-9262-z.

Sember, A., Bohlen, J., Šlechtová, V., Altmanová, M., Symonová, R., and Ráb, P. (2015). Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. *BMC Evol Biol* 15, 251. doi: 10.1186/s12862-015-0532-9.

Šíchová, J., Nguyen, P., Dalíková, M., and Marec, F. (2013). Chromosomal evolution in tortricid moths: conserved karyotypes with diverged features. *PLoS ONE* 8, e64520. doi: 10.1371/journal.pone.0064520.

Šíchová, J., Ohno, M., Dincă, V., Watanabe, M., Sahara, K., and Marec, F. (2016). Fissions, fusions, and translocations shaped the karyotype and multiple sex chromosome constitution of the northeast-Asian wood white butterfly, *Leptidea amurensis*. *Biol J Linn Soc* 118, 457–471. doi: 10.1111/bij.12756.

Šíchová, J., Voleníková, A., Dincă, V., Nguyen, P., Vila, R., Sahara, K., et al. (2015). Dynamic karyotype evolution and unique sex determination systems in *Leptidea* wood white butterflies. *BMC Evol Biol* 15, 89. doi: 10.1186/s12862-015-0375-4.

Singh, K. S., Troczka, B. J., Duarte, A., Balabanidou, V., Trissi, N., Carabajal Paladino, L. Z., et al. (2020). The genetic architecture of a host shift: an adaptive walk protected an aphid and its endosymbiont from plant chemical defenses. *Sci Adv* 6, eaba1070. doi: 10.1126/sciadv.aba1070.

Široký, J., Lysák, M. A., Doležel, J., Kejnovský, E., and Vyskot, B. (2001). Heterogenity of rDNA distribution and genome size in Silene spp. *Chromosome Res* 9, 387–393. doi: 10.1023/A:1016783501674.

Smith, D. A. S., Gordon, I. J., Traut, W., Herren, J., Collins, S., Martins, D. J., et al. (2016). A neo-W chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. *Proc R Soc B* 283, 20160821. doi: 10.1098/rspb.2016.0821.

Sochorová, J., Garcia, S., Gálvez, F., Symonová, R., and Kovařík, A. (2018). Evolutionary trends in animal ribosomal DNA loci: introduction to a new online database. *Chromosoma* 127, 141–150. doi: 10.1007/s00412-017-0651-8.

Sparks, M. E., Hebert, F. O., Johnston, J. S., Hamelin, R. C., Cusson, M., Levesque, R. C., et al. (2021). Sequencing, assembly and annotation of the whole-insect genome of *Lymantria dispar dispar*, the European gypsy moth. *G3-Genes Genom Genet* 11, jkab150. doi: 10.1093/g3journal/jkab150.

Stevens, T. J., Lando, D., Basu, S., Atkinson, L. P., Cao, Y., Lee, S. F., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544, 59–64. doi: 10.1038/nature21429.

Steward, R. A., Okamura, Y., Boggs, C. L., Vogel, H., and Wheat, C. W. (2021). The genome of the margined white butterfly (*Pieris macdunnoughii*): sex chromosome insights and the power of polishing with PoolSeq data. *Genome Biol Evol* 13, evab053. doi: 10.1093/gbe/evab053.

Suetsugu, Y., Futahashi, R., Kanamori, H., Kadono-Okuda, K., Sasanuma, S., Narukawa, J., et al. (2013). Large scale full-length cDNA sequencing reveals a unique genomic landscape in a

lepidopteran model insect, *Bombyx mori*. *G3-Genes Genom Genet* 3, 1481–1492. doi: 10.1534/g3.113.006239.

Takehana, Y., Hamaguchi, S., and Sakaizumi, M. (2008). Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, *Oryzias javanicus* and *O. hubbsi*. *Chromosome Res* 16, 801–811. doi: 10.1007/s10577-008-1227-5.

Talavera, G., Lukhtanov, V. A., Pierce, N. E., and Vila, R. (2013). Establishing criteria for higher-level classification using molecular data: the systematics of *Polyommatus* blue butterflies (Lepidoptera, Lycaenidae). *Cladistics* 29, 166–192. doi: 10.1111/j.1096-0031.2012.00421.x.

The Tree of Sex Consortium (2014). Tree of Sex: A database of sexual systems. *Sci Data* 1, 140015. doi: 10.1038/sdata.2014.15.

Traut, W., and Marec, F. (1996). Sex chromatin in Lepidoptera. *Q Rev Biol* 71, 239–256. doi: 10.1086/419371.

Traut, W., Sahara, K., and Marec, F. (2007). Sex chromosomes and sex determination in Lepidoptera. *Sex Dev* 1, 332–346. doi: 10.1159/000111765.

Twyford, A. D., and Friedman, J. (2015). Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion: genetics of adaptive divergence in Monkey flower. *Evolution* 69, 1476–1486. doi: 10.1111/evo.12663.

van Leen, E., Brückner, L., and Henssen, A. G. (2022). The genomic and spatial mobility of extrachromosomal DNA and its implications for cancer therapy. *Nat Genet* 54, 107–114. doi: 10.1038/s41588-021-01000-z.

Van Nieukerken, E. J., Kaila, L., Kitching, I. J., Kristensen, N. P., Lees, D. C., Minet, J., et al. (2011). Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa* 3148, 212. doi: 10.11646/zootaxa.3148.1.41.

van't Hof, A. E., Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., et al. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534, 102–105. doi: 10.1038/nature17951.

Van't Hof, A. E., Nguyen, P., Dalíková, M., Edmonds, N., Marec, F., and Saccheri, I. J. (2013). Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a model of industrial melanism. *Heredity* 110, 283–295. doi: 10.1038/hdy.2012.84.

Vara, C., Paytuví-Gallart, A., Cuartero, Y., Álvarez-González, L., Marín-Gual, L., Garcia, F., et al. (2021). The impact of chromosomal fusions on 3D genome folding and recombination in the germ line. *Nat Commun* 12, 2981. doi: 10.1038/s41467-021-23270-1.

Veyrunes, F., Waters, P. D., Miethke, P., Rens, W., McMillan, D., Alsop, A. E., et al. (2008). Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res* 18, 965–973. doi: 10.1101/gr.7101908.

Vítková, M., Fuková, I., Kubíčková, S., and Marec, F. (2007). Molecular divergence of the W chromosomes in pyralid moths (Lepidoptera). *Chromosome Res.* 15, 917–930. doi: 10.1007/s10577-007-1173-7.

Vítková, M., Král, J., Traut, W., Zrzavý, J., and Marec, F. (2005). The evolutionary origin of insect telomeric repeats, (TTAGG) $_N$. *Chromosome Res* 13, 145–156. doi: 10.1007/s10577-005-7721-0.

Voleníková, A. (2015) Karyotype and sex chromosomes analysis of two species from basal lepidopteran family Hepialidae (Lepidoptera: Hepialidae), Mgr. Thesis, in Czech. 58 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Vondrak, T., Oliveira, L., Novák, P., Koblížková, A., Neumann, P., and Macas, J. (2021). Complex sequence organization of heterochromatin in the holocentric plant *Cuscuta europaea* elucidated by the computational analysis of nanopore reads. *Comput Struct Biotechnol* 19, 2179–2189. doi: 10.1016/j.csbj.2021.04.011.

Wallace, A. R. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. III. On the tendency of varieties to depart indefinitely from the original type. *J Proc Linn Soc London* 3, 53–62.

Wan, F., Yin, C., Tang, R., Chen, M., Wu, Q., Huang, C., et al. (2019). A chromosome-level genome assembly of *Cydia pomonella* provides insights into chemical ecology and insecticide resistance. *Nat Commun* 10, 4237. doi: 10.1038/s41467-019-12175-9.

Ward, C. M., Perry, K. D., Baker, G., Powis, K., Heckel, D. G., and Baxter, S. W. (2021). A haploid diamondback moth (*Plutella xylostella* L.) genome assembly resolves 31 chromosomes and identifies a diamide resistance mutation. *Insect Biochem Mol Biol* 138, 103622. doi: 10.1016/j.ibmb.2021.103622.

Wellband, K., Mérot, C., Linnansaari, T., Elliott, J. A. K., Curry, R. A., and Bernatchez, L. (2019). Chromosomal fusion and life history-associated genomic variation contribute to within-river local adaptation of Atlantic salmon. *Mol Ecol* 28, 1439–1459. doi: 10.1111/mec.14965.

Wilson Sayres, M. A. (2018). Genetic diversity on the sex chromosomes. *Genome Biol Evol* 10, 1064–1078. doi: 10.1093/gbe/evy039.

Wolf, K. W., Novak, K., and Marec, F. (1997). Kinetic organization of metaphase I bivalents in spermatogenesis of Lepidoptera and Trichoptera species with small chromosome numbers. *Heredity* 79, 35—143.

Wright, A. E., Dean, R., Zimmer, F., and Mank, J. E. (2016). How to make a sex chromosome. *Nat Commun* 7, 12087. doi: 10.1038/ncomms12087.

Wurster, D. H., and Benirschke, K. (1970). Indian muntjac, *Muntiacus muntjak*: a deer with a low diploid chromosome number. *Science* 168, 1364–1366. doi: 10.1126/science.168.3937.1364.

Xia, Q., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., et al. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* 306, 1937–1940. doi: 10.1126/science.1102210.

Yamamoto, K., Nohata, J., Kadono-Okuda, K., Narukawa, J., Sasanuma, M., Sasanuma, S., et al. (2008). A BAC-based integrated linkage map of the silkworm *Bombyx mori*. *Genome Biol* 9, R21. doi: 10.1186/gb-2008-9-1-r21.

Yano, C. F., Merlo, M. A., Portela-Bens, S., Cioffi, M. de B., Bertollo, L. A. C., Santos-Júnior, C. D., et al. (2020). Evolutionary dynamics of multigene families in *Triportheus* (Characiformes,

Triportheidae): A transposon mediated mechanism? *Front Mar Sci* 7. doi: 10.3389/fmars.2020.00006.

Yasukochi, Y., Fujii, H., and Goldsmith, M. R. (2008). "Silkworm.," in *Genome mapping and genomics in Arthropods* (Berlin, Heidelberg: Springer Berlin Heidelberg), 43–57. Available at: http://link.springer.com/10.1007/978-3-540-73833-6_4 [Accessed June 8, 2022].

Yasukochi, Y., Ohno, M., Shibata, F., Jouraku, A., Nakano, R., Ishikawa, Y., et al. (2016). A FISH-based chromosome map for the European corn borer yields insights into ancient chromosomal fusions in the silkworm. *Heredity* 116, 75–83. doi: 10.1038/hdy.2015.72.

Yasukochi, Y., Tanaka-Okuyama, M., Shibata, F., Yoshido, A., Marec, F., Wu, C., et al. (2009). Extensive conserved synteny of genes between the karyotypes of *Manduca sexta* and *Bombyx mori* revealed by BAC-FISH mapping. *PLoS ONE* 4, e7465. doi: 10.1371/journal.pone.0007465.

Yoshido, A., Bando, H., Yasukochi, Y., and Sahara, K. (2005a). The *Bombyx mori* karyotype and the assignment of linkage groups. *Genetics* 170, 675–685. doi: 10.1534/genetics.104.040352.

Yoshido, A., Marec, F., and Sahara, K. (2005b). Resolution of sex chromosome constitution by genomic in situ hybridization and fluorescence in situ hybridization with (TTAGG)n telomeric probe in some species of Lepidoptera. *Chromosoma* 114, 193–202.

Yoshido, A., Šíchová, J., Pospíšilová, K., Nguyen, P., Voleníková, A., Šafář, J., et al. (2020). Evolution of multiple sex-chromosomes associated with dynamic genome reshuffling in Leptidea wood-white butterflies. *Heredity* 125, 138–154. doi: 10.1038/s41437-020-0325-9.

Yoshido, A., Yasukochi, Y., and Sahara, K. (2011). *Samia cynthia* versus *Bombyx mori*: comparative gene mapping between a species with a low-number karyotype and the model species of Lepidoptera. *Insect Biochem Mol Biol* 41, 370–377. doi: 10.1016/j.ibmb.2011.02.005.

You, M., Yue, Z., He, W., Yang, X., Yang, G., Xie, M., et al. (2013). A heterozygous moth genome provides insights into herbivory and detoxification. *Nat Genet* 45, 220–225. doi: 10.1038/ng.2524.

Yusuf, M., Sajid, A., Robinson, I. K., and Lalani, E.-N. (2022). 3D ultrastructural imaging of chromosomes using serial block-face scanning electron microscopy (SBFSEM). *DNA* 2, 30–43. doi: 10.3390/dna2010003.

Zattera, M. L., Gazolla, C. B., Soares, A. de A., Gazoni, T., Pollet, N., Recco-Pimentel, S. M., et al. (2020). Evolutionary dynamics of the repetitive DNA in the karyotypes of *Pipa carvalhoi* and *Xenopus tropicalis* (Anura, Pipidae). *Front Genet* 11, 637. doi: 10.3389/fgene.2020.00637.

Zhang, X., and Wang, T. (2021). Plant 3D chromatin organization: important insights from chromosome conformation capture analyses of the last 10 years. *Plant Cell Physiol* 62, 1648–1661. doi: 10.1093/pcp/pcab134.

Zhou, Q., and Bachtrog, D. (2012). Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science* 337, 341–345. doi: 10.1126/science.1225385.

Zhou, Y., Shearwin-Whyatt, L., Li, J., Song, Z., Hayakawa, T., Stevens, D., et al. (2021). Platypus and echidna genomes reveal mammalian biology and evolution. *Nature* 592, 756–762. doi: 10.1038/s41586-020-03039-0.

6. Curriculum vitae

Mgr. Irena Provazníková

Faculty of Science, University of South Bohemia in České Budějovice, Branišovská 1760/31a, 370 05 České Budějovice, Czech Republic

PERSONAL INFORMATION

Maiden name: Hladová e-mail: irenka.provaznikova@gmail.com

Date of birth: November 22 1989

Place of birth: Plzeň, Czech Republic

Nationality: Czech

RESEARCH INTERESTS

Karyotype and sex chromosome evolution in Lepidoptera, molecular cytogenetics, comparative genomics in non-model organisms, phylogenomics, genetics of ecological adaptation and speciation

---

EDUCATION

2016 – present	Ph.D. student in Molecular and Cell Biology and Genetics, Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic
2013 – 2016	Master's degree in Experimental Biology – Genetics and Gene Engineering, Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic
2009 – 2013	Bachelor's degree in Biology, Faculty of Science University of South Bohemia in České Budějovice, Czech Republic

THESES

Dissertation thesis	Role of chromosomal rearrangements in evolution of Lepidoptera - in progress
Master thesis	New cytogenetic markers and evolutional dynamic of karyotypes in Lepidoptera
Bachelor thesis	SNP identification for genetic mapping of CpGV resistance in codling moth, *Cydia pomonella*

RESEARCH AND WORK EXPERIENCE

2019 – present	Training lab technician, EMBL's International Centre for Advanced Training (EICAT), EMBL, Heidelberg, Germany
2016 – 2019	Member of Laboratory of Cytogenetics and Comparative Genomics, Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic

2011 – 2019 Member of Laboratory of Molecular Cytogenetics, Institute of Entomology, Biology Centre AS CR, České Budějovice

INTERNATIONAL EXPERIENCE, COURSES AND SCHOLARSHIPS

2019 8th RepeatExplorer Workshop on the Application of Next Generation Sequencing to Repetitive DNA Analysis, Institute of Plant Molecular Biology, České Budějovice, Czech Republic (21-23 May)

2018 internship in the Institute of Parasitology SAS, Košice, Slovakia (1 week)

2017 Erasmus+ EU Program scholarship, research internship, Lund University, Sweden (3 months)

2016-2017 internship in Genomics Core Facility, EMBL, Heidelberg, Germany (4 weeks)

2014 internship in Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina (1 month)

RESEARCH GRANTS

2014 Student Grant Agency of Faculty of Science, University of South Bohemia in České Budějovice grant: *CpGV bioassay on C. pomonella*

AWARDS

2016 Rector's award for outstanding research results gained in Master thesis, Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic

2014 Outstanding academic achievements, Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic (12 months stipendium)

PUBLICATIONS

Voleníková A., Nguyen P., Davey P., Sehadová H., Kludkiewicz B., Koutecký P., Walters J.R., Roessingh P., **Provazníková I.**, Šerý M., Žurovcová M., Hradilová M., Rouhová L., Žurovec M. (2022) Genome sequence and silkomics of the spindle ermine moth, *Yponomeuta cagnagella*, representing the early-diverging lineage of the ditrysian Lepidoptera. *Commun Biol* 5: 1281 DOI: 10.1038/s42003-022-04240-9

Dalíková M., **Provazníková I.**, Provazník J., Grof-Tisza P., Pepi A., Nguyen P. (under revision in Genome Biol Evol) The role of repetitive DNA in re-patterning of major rDNA clusters in Lepidoptera. bioRxiv (2022) DOI: 10.1101/2022.03.26.485928

Provazníková I., Hejníčková M., Visser S., Dalíková M., Carabajal-Paladino L.Z., Zrzavá M., Voleníková A., Marec F., Nguzen P. (2021) Large-scale comparative analysis of cytogenetic markers across Lepidoptera. *Scietific Reports* 11: 12214 DOI: 10.1038/s41598-021-91665-7

Ferguson K.B., Visser S., Dalíková M., **Provazníková I.**, Urbaneja A., Pérez-Hedo M., Marec F., Werren J.H., Zwaan B.J., Pannebakker B.A., Verhulst E.C. (2021) Jekyll or Hyde? The genome (and

more) of Nesidiocoris tenuis, a zoophytophagous predatory bug that is both a biological control agent and a pest. Insect Molecular Biology, 30: 188-209. DOI: 10.1111/imb.12688

Orosová M., **Provazníková I.**, Bin W.X., Oros M. (2019) Chromosomal study of *Khawia abbottinae* (Cestoda: Caryophyllidea): karyotype and localization of telomeric and ribosomal sequences after fluorescence in situ hybridization (FISH). *Parasitology Research* 118(10):2789-2800. DOI: 10.1007/s00436-019-06450-3.

Carabajal Paladino L.Z., **Provazníková I.**, Berger M., Bass C., Aratchige N. S., López S.N., Marec F., Petr Nguyen P. (2019) Sex chromosome turnover in moths of the diverse superfamily Gelechioidea. *Genome Biology and Evolution* 11 : 1307–1319. DOI: 10.1093/gbe/evz075

Zrzavá M., **Hladová I.**, Dalíková M., Šíchová J., Õunap E., Kubíčková S., Marec F. (2018) Sex chromosomes of the iconic moth *Abraxas grossulariata* (Lepidoptera, Geometridae) and its congener *A. sylvata*. *Genes* 9 : 279. DOI: 10.3390/genes9060279

Vazač J., Füssy Z., **Hladová I.,** Killi S., Oborník M. (2018) Ploidy and Number of Chromosomes in the Alveolate Alga *Chromera velia*. *Protist* 169 : 53-63. DOI: 10.1016/j.protis.2017.12.001

Dalíková M., Zrzavá M., **Hladová I.**, Nguyen P., Šonský I., Flegrová M., Kubíčková S., Voleníková A., Kawahara A.Y., Peters R.S., Marec F. (2017) New insights into the evolution of the W chromosome in Lepidoptera. *Journal of Heredity* 108 : 709-719. DOI: 10.1093/jhered/esx063

CONFERENCE CONTRIBUTIONS (selected)

Provazníková I., Dalíková M., Provazník J., Nguyen P. (2021) The potential role of transposable elements in evolution of major rDNA clusters in Lepidoptera. The Mobile Genome: Genetic and Physiological Impacts of Transposable Elements (virtual). August 29 September 1, 2021. EMBL Heidelberg, Germany. (poster)

Hladová I., Roessingh P., Provazník J., Šlajsová M., Sahara K., Marec F., Nguyen P. (2018) Karyotype and sex chromosome evolution of small ermine moths of the genus *Yponomeuta*. 10th International Workshop on the Molecular Biology and Genetics of the Lepidoptera. August 19-25, 2018. Orthodox Academy of Crete, Greece. (oral presentation)

Hladová I., Voleníková A., Marec F., Carabajal Paladino L. Z., Nguyen P. (2014) Karyotype evolution in Lepidoptera: Insight from basal lineages. The 9th International Workshop on the Molecular Biology and Genetics of the Lepidoptera, 17-23 August 2014, Orthodox Academy of Crete, Kolympari, Crete. (poster)

TEACHING (selected)

2016 – 2019	Mentoring of Master students (co-supervisor)
2016 – 2019	"Practical Course in Genetics“ course for university students, Faculty of Science, University of South Bohemia in České Budějovice (lector)
2014 – 2019	"Cytogenetics - laboratories" course for university students at Faculty of Science, University of South Bohemia in České Budějovice (lector)

irena.provaznikova@embl.de

Drivers of karyotype evolution in Lepidoptera
Ph.D. Thesis, 2022

University of South Bohemia in České Budějovice
Faculty of Science
Branišovská 1760
CZ-37005 České Budějovice, Czech Republic

Phone: +420 387 776 201
www.prf.jcu.cz, e-mail: sekret-fpr@prf.jcu.cz